Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Using Bayesian Mixing Models to Unravel Isotopic Data and Quantify N_2O Production and Consumption Pathways

Master Thesis

Philipp Fischer
fischphi@student.ethz.ch

Swiss Data Science Center
ETH Zürich

Supervisors:
Prof. Dr. Fernando Perez Cruz
PD Dr. Eliza Harris
PD Dr. Dominika Lewicka-Szczebak
Dr. Maciej Lewicki

June 2, 2023

I would like to thank Fernando Perez Cruz for giving me the opportunity to write this thesis at the Swiss Data Science Center and for being open and available for requests at any time throughout my work.

I am extremely grateful to Eliza Harris for supervising this thesis and supporting me with advice, guidance, inspiration and keen ideas. Her humble yet resolute approach to projects like this is truly commendable and I am sure that without her help this thesis would not have been possible. My sincere thanks go to Dominika Lewicka-Szczebak and Maciej Lewicki, who also supervised this thesis, leveraging their expertise from previous work to give me a head start into the area of isotopic data analysis, that was previously unfamiliar to me.

Furthermore, I am thankful to Matti Barthel, Charlotte Decock and William Aeberhard who all took time to give advice and discuss various subjects with me. Isotopic data is hard to come by and so I am deeply grateful to Eliza Harris, Elena Stoll, Matti Barthel, Charlotte Decock, Elizabeth Verhoeven and Nora Galarotti for sharing their datasets with me.

Finally, my special thanks go to my friends and family who relentlessly supported and encouraged me, and whose well-timed distractions proved to be invaluable throughout this journey.

Abstract

Isotopic measurements of N_2O contain valuable information about its production and consumption pathways, whose rigorous understanding could shed light on currently poorly understood processes, such as the recent acceleration of N_2O build-up in the atmosphere and its connection to anthropogenic activities. Contributions to isotopic mixtures can be quantified by solving systems of mixing equations, although this introduces mathematical constraints and complicates handling uncertainties. Bayesian hierarchical models combine the use of expert information with measured data and a mathematical mixing model while considering and updating the uncertainties involved. Models specialized to solve mixing problems are known as Bayesian mixing models and have already been successful in generating insights from isotopic data in many different contexts. However, data analysis of N_2O production and consumption pathways requires managing time series data and simultaneous estimation of endmember fractionation, which has not been combined in a modeling approach before.

In total, four different model classes are presented in this thesis: independent time step models, Gaussian process priors on measurements, Dirichlet-Gaussian process priors, and generalized linear models with spline bases. All four have been extensively tested in different variations and for a multitude of scenarios. Dirichlet-Gaussian process prior models have been found to be most reliable, allowing for simultaneous estimation of hyperparameters via Bayesian hierarchical modeling. Generalized linear models with spline bases seem promising as well, especially for fractionation estimation, although the robustness to real datasets is difficult to assess given their high flexibility. Additional experiments concluded that model performance of all classes could greatly be increased by reducing uncertainty in input data, whereas the addition of more isotopic measurements yielded a comparatively small benefit.

The `TimeFRAME` package available in `R` was implemented to distribute these models using Stan for parameter estimation in addition to supplementary functions re-implementing some of the surveyed isotope analysis techniques.

Contents

Chapter 1

Introduction

1.1. Background

Analysis of isotopic signatures is extensively used in geosciences and ecology as a way of studying mixing processes that are difficult or impossible to measure directly. Notable examples include the study of water contamination [1], animal diet composition [2] and most importantly for this thesis the quantification of nitrogen pathways [3], [4].

Isotopic signature measurements of nitrous oxide (N_2O) have been shown to be useful for the understanding of its production and consumption pathways. Moreover, the analysis techniques used are applicable to many different research questions involving isotopic data [5], [6]. Nitrous oxide is a naturally occurring gas that is primarily produced by microbes in soil and water. It appears as a side-product as well as an intermediary in the series of processes constituting the nitrogen cycle. This cycle, and thus by proxy the production of N_2O, is impacted by anthropogenic activities such as fertilization [7]. However, the acceleration of its emission rate in the past few decades cannot be explained by rising fertilizer use alone, but could follow from the greater impact of climate change [8] or redistribution of nitrogen inputs [4]. This is critical since N_2O acts as a strong greenhouse gas and ozone depletant in the atmosphere [9], [10]. Quantification of individual production and consumption pathways despite complex interactions of numerous biotic and abiotic processes can be helpful to gain a better understanding of the causes and preventative measures regarding this development.

1.1.1. Nitrogen Cycle

The nitrogen cycle is commonly split into two main pathways. *Nitrification* is the oxidation of ammonium (NH_4^+) to nitrite (NO_2^-) and later nitrate (NO_3^-) performed by bacteria and other microorganisms [11]. It produces nitrous oxide (N_2O) as a side-product and has generally a low throughput compared to the rest of the nitrogen cycle. *Denitrification* is a process in the opposing direction, namely the reduction of nitrate to elemental nitrogen (N_2) having nitrous oxide as the penultimate intermediate. A multitude of organisms can be identified to participate in the denitrification process, since it can serve as a source of energy for them [12]. Both pathways are amplified when fertilization increases the available nitrogen in soil.

The different main pathways and even different mechanisms facilitating each individual pathway have distinct reaction properties which can be measured and compared using the isotopic signature of the corresponding product. This allows one to distinguish among them and even quantify individual contributions given sufficient measurement accuracy and advanced calculation methods.

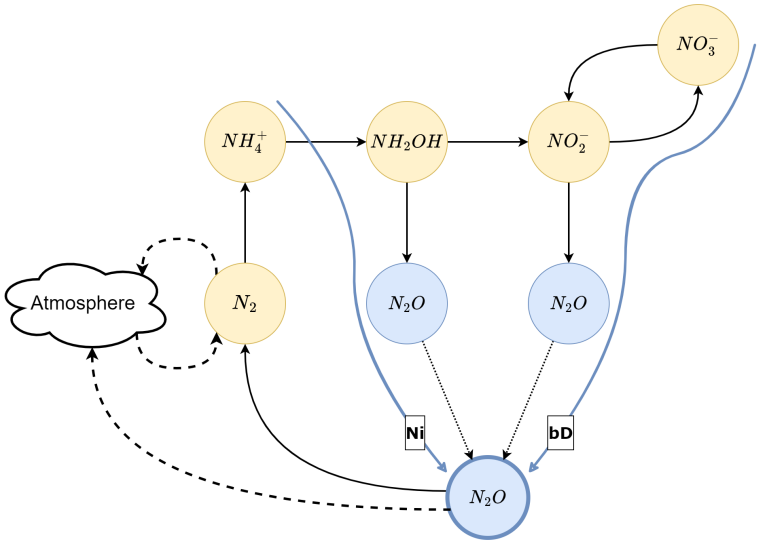

Figure 1.1.: Graphical illustration of the different stages in the nitrogen cycle. Importantly, nitrous oxide is produced via two main pathways, namely nitrification (Ni) and bacterial denitrification (bD). Both nitrous oxide and elemental nitrogen are typically in exchange with the atmosphere making this a partially open system.

1.1.2. Isotopic Signatures

Chemical elements are determined by the proton count in the nucleus of the atom. Isotopes are variations of the same element with a different number of neutrons, which are neutral particles also found within the nucleus. Molecules that differ in the isotopic composition of their atoms but are otherwise identical are known as isotopocules. They have almost identical chemical properties but vary slightly in their mass, which can have an impact on the rate of reactions.

Nitrous oxide consists of nitrogen with its most abundant isotope having a total of 14 nucleons ^{14}N and oxygen with the most common nucleon count being 16 thus denoting it ^{16}O. Their second most frequently occurring isotopes are ^{15}N with a natural abundance of 0.36% and ^{18}O with an abundance of 0.2% [13]. Possible substitutions of the most abundant nitrous oxide isotopocule ^{14}N–^{14}N–^{16}O are thus central nitrogen substitution ^{14}N–^{15}N–^{16}O known as α-configuration, terminal nitrogen substitution ^{15}N–^{14}N–^{16}O known as β-configuration and oxygen substitution ^{14}N–^{14}N–^{18}O. Multiple substitutions, known as clumped isotopes, are simultaneously possible, but given the low abundance of either heavy isotope they are exceptionally rare.

Example measurements of the isotopic signature of nitrous oxide would be the presence of nitrogen substitution δ^{15}N, the presence of oxygen substitution δ^{18}O and the difference between the α and β position known as site preference $\delta^{15}N^{SP}$. Depending on the source and particular reaction, a given substance such as nitrous oxide shows a characteristic isotopic signature and hence acts as a natural tracer for the specific source and mechanism.

1.2. Previous Work

Isotopic measurements have been used for various studies of nitrogen pathways, whether it be the main cycle of nitrification and denitrification or alternative processes that exist in conjunction with them. Typically, the endmember isotopic signature is partitioned into the different sources and pathways by constructing a linear system of equations and solving for the source contributions. This technique is known as isotopic mapping [14], [15]. Analysis using statistical frameworks provides an alternative that is much more open to different measurement designs and can put produced estimates in context of the associated uncertainties [16].

Since it has been usual to measure only two types of isotopic measurements for nitrous oxide in the past, linear isotope mapping has been straightforward to compute and simple enough to extend with rudimentary ways of handling uncertainty. With the recent possibility of measuring more than two isotopic properties of the nitrous oxide molecule, more sophisticated methods could be developed and applied. The *3DIM* model [17] uses all three above mentioned isotopic measurements, $\delta^{15}\mathrm{N}$, $\delta^{15}\mathrm{N}^{\mathrm{SP}}$ as well as $\delta^{18}\mathrm{O}$, and places them in a Bayesian framework to account for mixing of multiple sources as well as the further reduction to elemental nitrogen under uncertainty.

The recently published *FRAME* model [18] provides a Bayesian framework for a wide range of isotopic measurements with different cases of production and consumption pathways and is accompanied by a graphical interface implemented in Python [19]. It casts a wide net in terms of the problems it can address by allowing arbitrary mixing equations augmented with auxiliary variables to model the measurements of the endmember as the combination of many different contributing processes.

1.3. Aim and Motivation

Although at the present time there exist models with implementations that deal with isotopic measurements for the purpose of estimating source contributions [17], [18], [20]–[22], they appear to be tailored toward isotopic problems that do not deal with fractionation or have no straightforward way of benefiting from time series information.

The goal of this thesis is thus to provide a statistical framework as well as a corresponding implementation that can be used to quantitatively analyze isotopic mixtures over time. In particular, the objective is to improve and extend existing Bayesian models to deal with multiple isotopic measurements in the presence of fractionation and to meaningfully incorporate time series information. Models shall be compared in terms of accuracy of estimates as well as time series properties such as the rate of change on simulated data sets.

Chapter 2

Theory

2.1. Stable Isotope Mixing

In order to study and quantify different sources and pathways of a particular chemical species, stable isotope measurements can be used as indicators. Fundamental notation as well as some calculations shall be established in the following section based on Hayes [23].

In the most general case the measurements yield concentrations for the isotope of interest denoted by C against the concentration of the most abundant, lighter reference isotope C_0. For instance in nitrogen pathway analysis the presence of the heavier ^{15}N is measured in contrast to the most abundant stable isotope ^{14}N. Calculations are then based on the *isotope ratio* defined by the ratio of concentrations.

$$R := \frac{C}{C_0} \tag{2.1}$$

As is the case for the vast majority of isotopic analyses of light elements it can be assumed that the most abundant isotope has a much higher concentration $C_0 \gg C$ and therefore the isotope ratio is small. For the purpose of calibration to an international standard scale a different notation is commonly used, where the isotope ratio is expressed relative to a standardized ratio R_{std} designated and agreed upon on a case by case basis [24]. Nitrogen is standardized using the isotopic ratio of elementary nitrogen in air (N2-AIR).

$$\delta := \frac{R}{R_{std}} - 1 \tag{2.2}$$

Additionally, the values are typically expressed in permill (‰) to make them more readable and usable. In this delta notation the isotope of interest is appended to the symbol to uniquely identify the measurement that is being made. In the case of nitrogen the bulk isotope measurement is therefore denoted by $\delta^{15}N$.

Isotopic mixing calculations are used when the measured species is known to be the mixture of two or more sources with different isotopic compositions. The sources shall be enumerated by $k = 1, \ldots, K$ with each having concentration C_k and thus isotope ratio R_k. Given that the sources are mixed with molar proportions $f_1, \ldots, f_K$ having $\sum_{k=1}^{K} f_k = 1$, the mass balance equation for the mixture can be derived in terms of the concentration values by normalizing with the total sum of constituents.

$$\frac{C}{C_0 + C} = \sum_{k=1}^{K} \frac{f_k C_k}{C_{0,k} + C_k} \tag{2.3}$$

One can easily see that using the assumption of a very large light isotope concentration $C_0 \gg C$ the balance equation holds for the isotope ratios as well $R = \sum_{k=1}^{K} f_k R_k$ and furthermore, using the assumption that the ratios do not deviate much from the standard ratio R_{std}, the equation even holds approximately for delta values with $\delta_k = \frac{R_k}{R_{std}} - 1$.

$$\delta \approx \sum_{k=1}^{K} f_k \delta_k \tag{2.4}$$

2.1.1. Fractionation Effects

One might want to take into account that a source in this case is not simply a collection of pre-existing product, but a substrate that undergoes a chemical reaction to become said product. Even if the original isotopic ratio of the source substrate corresponding to source k is given by $\tilde{R}_k$ the immediate products of the reaction are not necessarily isotopically identical. For instance the isotopic signature of nitrite in soil will be different from nitrous oxide purely produced from denitrification. This is due to the fact that chemical reactions do not preserve the isotopic ratios, but they generally favor certain isotopes and thereby introduce a so called *fractionation effect*. A typical example is the fact that light isotopes tend to react faster and thus products get enriched in light isotopes compared to their substrates. This effect is of course only relevant if the isotopes can separate in the first place and the reaction is not fully complete, otherwise the exact same isotopic ratios would necessarily be observed due to mass balance.

Fractionation effects are quantified by the relation of the isotopic ratios of substrates and products. The so called *fractionation factor* is defined as the ratio of isotope-specific reaction rate constants and consequently corresponds to the relationship of product and substrate isotopic ratio $\alpha_k = \frac{R_k}{\tilde{R}_k}$ [25]. Inheriting the notation convention from delta values, it is similarly expressed in permill (‰).

$$\varepsilon_k := \frac{R_k}{\tilde{R}_k} - 1 \tag{2.5}$$

This allows the expression of the mixture isotope ratio as a weighted sum of the original substrate isotope ratios and fractionation effects, which by approximation on small deviations relative to the standard isotope ratio also holds in delta notation, as already seen above.

$$R = \sum_{k=1}^{K} f_k \alpha_k \tilde{R}_k \tag{2.6}$$

In addition to the sources producing the chemical species of interest, one might also consider reactions that consume it and thus induce a fractionation effect of the opposite size. For instance reactions that favor light isotopes will remove them and thus leave the mixture enriched in heavy isotopes compared to what was produced by the different sources. This is typically the case for the reduction of nitrous oxide to elemental nitrogen.

Under the assumption that the substrate has isotope ratio $\tilde{R}$ directly after mixing and is then transformed into a product with isotope ratio $R_p = \alpha R$ determined by the fractionation factor ε of the reaction, the resulting isotopic ratio can be derived. Using an approximation of small

values of ε this relationship is stated in delta notation as $\delta_p = \delta + \varepsilon$. A common situation involves open systems where reactants can be supplied infinitely and the transformed product is instantly lost to the environment. If such systems are in steady state the mass balance equation can be interpreted per unit time and adapted to delta notation in terms of the fraction of remaining substrate $r \in [0, 1]$, also called *fractionation weight* or *fractionation index*.

$$\tilde{\delta} = r\delta + (1 - r)\delta_p = r\delta + (1 - r)(\delta + \varepsilon) \tag{2.7}$$

Now the terms can be reordered and the previous sources can be incorporated.

$$\delta = \tilde{\delta} - \varepsilon(1 - r) = \sum_{k=1}^{K} f_k \delta_k - \varepsilon(1 - r) \tag{2.8}$$

A fully closed system has limited resources of reactants in addition to the accumulation of product such that mass balance is always satisfied. This fractionation type results in reaction kinetics described by the Rayleigh equations which were developed by John William Strutt and is therefore oftentimes referred to as *Rayleigh fractionation*. The Rayleigh equation then states the relationship of the original isotope ratio $\tilde{R}$ and the isotope ratio of the remaining substrate R in terms of its molar proportion r.

$$\frac{R}{\tilde{R}} = r^{\varepsilon} \tag{2.9}$$

Restating this equation in terms of the delta notation and using the approximation $\delta \approx \log(1 + \delta)$ for small values the full mixing equation including sources and product fractionation can be derived.

$$\delta = \tilde{\delta} + \varepsilon \log r = \sum_{k=1}^{K} f_k \delta_k + \varepsilon \log r \tag{2.10}$$

Careful consideration needs to be taken for the approximations involved, since for large yields $r \approx 0$ they can lead to errors that are of the same order of magnitude as the effects to be studied.

A different closed system interaction can occur when the reaction in question is reversible. The fractionation effects in both direction can then be used determine an isotopic signature at equilibrium δ_{eq}. If the system is not fully equilibrated and thus not in steady state the isotopic signature of the substrate can be expressed in terms of the equilibrated proportion $r_{eq} \in [0, 1]$.

$$\delta = (1 - r_{eq})\tilde{\delta} + r_{eq}\delta_{eq} = (1 - r_{eq}) \sum_{k=1}^{K} f_k \delta_k + r_{eq}\delta_{eq} \tag{2.11}$$

These three mixing and fractionation equations are approximations for open system in steady state as well as closed systems with and without reversible reactions. They are useful for basic analyses, but many additional cases exists with fractionation equations that are not covered here. Some examples are the exact formulations of the above equations [26] as well as individual source fractionation before mixing occurs [15]. Multiple fractionation terms in one equation are also possible, in which case the fractionation proportions will be enumerated by $r_1, \ldots, r_L$.

2.1.2. Inference of Source Contributions

One objective of studying isotopic signatures is to determine the source contributions $f_1, \dots, f_K$ from measurements of the mixture. However, measuring one single type of isotope will only be efficient in distinguishing between two sources. For additional sources or if consumption of the mixture needs to be taken into account, multiple isotopic measurements are necessary. Analysis of nitrous oxide sources and pathways for instance can include the bulk isotopic abundance of heavy nitrogen $\delta^{15}\mathrm{N}$ and heavy oxygen $\delta^{18}\mathrm{O}$ in addition to nitrogen site preference $\delta^{15}\mathrm{N}^{\mathrm{SP}}$.

In general the vector of d different isotopic measurements shall be denoted by $X \in \mathbb{R}^d$. Measurements of the isotopic signature of each individual source enumerated by $k = 1, \dots, K$ are assumed to be known and denoted by $S_1, \dots, S_k \in \mathbb{R}^d$ together with the fractionation factor $A \in \mathbb{R}^d$. Using vector and matrix notation they can later be used to state the mixing equation in vector form.

$$\mathbf{f} := [f_1 \cdots f_K]^{\mathrm{T}} \in \mathbb{R}^K$$
$$\mathbf{S} := [S_1 \dots S_K] \in \mathbb{R}^{d \times K}$$

The case of Rayleigh fractionation shall serve as a representative example for this section since the simple inference methods discussed here are tailored to this specific example. More sophisticated methods are developed for the general case and the mixing equation $\mu(\mathbf{f}, r)$ can be expressed in terms of $\mathbf{f}$ and r as is deemed appropriate for the given situation. The example mixing and fractionation equation using Rayleigh fractionation can be expressed in vectorized form as well.

$$X = \mu(\mathbf{f}, r) := \sum_{k=1}^{K} f_k S_k + A \log r = \mathbf{S}\mathbf{f} + A \log r \tag{2.12}$$

Given that the number of measurements d is chosen exactly such that the source contributions $\mathbf{f}$ can be identified uniquely, a method of determining them is the isotope mapping approach [15]. For the simplest example of only two sources and two measurements $K = d = 2$ the mixing equation can be solved for the parameters of interest.

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} S_{11} \\ S_{12} \end{bmatrix} f_1 + \begin{bmatrix} S_{21} \\ S_{22} \end{bmatrix} f_2 + \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \log r = \begin{bmatrix} S_{21} \\ S_{22} \end{bmatrix} + \begin{bmatrix} S_{11} - S_{21} \\ S_{12} - S_{22} \end{bmatrix} f_1 + \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \log r$$

This is a linear system of equations that can be addressed with the usual techniques giving explicit solutions to compute the desired quantities.

$$\begin{aligned} f_1 = 1 - f_2 &= \frac{A_1(X_2 - S_{22}) - A_2(X_1 - S_{21})}{A_1(S_{12} - S_{22}) - A_2(S_{11} - S_{21})} \\ \log r &= \frac{S_{12}(X_1 - S_{21}) - S_{11}(X_2 - S_{22}) + X_2 S_{21} - X_1 S_{22}}{A_1(S_{12} - S_{22}) - A_2(S_{11} - S_{21})} \end{aligned} \tag{2.13}$$

No unique solution exists if the quantity in both denominators is equal to zero or equivalently if $\frac{A_1}{A_2} = \frac{S_{11} - S_{21}}{S_{12} - S_{22}}$. In this case the mixing line and the consumption line coincide and the effects of both can no longer be separated as the problem essentially reduces to fulfilling a one dimensional constraint with two unknowns.

The more general case of mixing and Rayleigh fractionation can be solved for an arbitrary number of sources as long as an equal number of isotopic measurements exist $K = d \geq 2$. In that case the linear system of equations can be written in matrix terms and augmented with the sum constraint on $\mathbf{f}$ to ensure that the solution fulfills it exactly.

$$\tilde{X} := \begin{bmatrix} X \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{S} & -A \\ \mathbf{1}^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ -\log r \end{bmatrix} =: \tilde{\mathbf{S}}\tilde{\mathbf{f}} \tag{2.14}$$

This d+1-dimensional linear system of equations can be addressed with decomposition techniques and its solution can be expressed as $\tilde{\mathbf{f}} = \tilde{\mathbf{S}}^{-1}\tilde{X}$. A unique solution exists if $\tilde{\mathbf{S}}$ is invertible or equivalently if none of the mixing lines as well as the consumption line are collinear. Only non-negative solutions $\tilde{\mathbf{f}} \geq 0$ are feasible to ensure that the source contributions $\mathbf{f}$ correspond to mixing weights and that $0 < r \leq 1$.

2.1.3. Uncertainty Quantification

A flaw of the isotope mapping approach as presented above is that it does not take measurement uncertainty into account. However, this can easily be added by formulating the measurements X as random variables with expected value given by the mixing equation $\mathbb{E}[X] = \mu(\mathbf{f}, r)$. Most commonly, measurements are modeled using the Gaussian distribution with independent components and variance $\eta^2 \in \mathbb{R}$.

$$X \sim \mathcal{N}_d\big(\mu(\mathbf{f}, r), \eta^2\mathbb{1}\big) = \mathcal{N}_d\big(\mathbf{S}\mathbf{f} + A\log r, \eta^2\mathbb{1}\big) \tag{2.15}$$

The likelihood equation can now be stated in terms of the multivariate Gaussian density with independent components and the maximization problem can be solved for $\mathbf{f}$ and r under consideration of all relevant constraints.

$$p(X|\mathbf{f}, r) = \left(2\pi\eta^2\right)^{-\frac{d}{2}} \exp\left(-\frac{\|X - \mu(\mathbf{f}, r)\|^2}{2\eta^2}\right)$$

Solving for the maximum likelihood solution thus leads to the identical solution derived from solving Equation 2.14 if it exists and otherwise the closest feasible solution in terms of the Euclidean distance by solving the optimization problem.

$$\underset{\mathbf{f}\in\mathcal{S}_K, r\in[0,1]}{\operatorname{argmax}} \; p(X|\mathbf{f}, r) = \underset{\mathbf{f}\in\mathcal{S}_K, r\in[0,1]}{\operatorname{argmin}} \; \|X - \mu(\mathbf{f}, r)\|^2 = \underset{\tilde{\mathbf{f}}\geq 0}{\operatorname{argmin}} \; \|\tilde{X} - \tilde{\mathbf{S}}\tilde{\mathbf{f}}\|^2 \tag{2.16}$$

The least squares minimizer can be expressed using the Moore-Penrose pseudo-inverse $\tilde{\mathbf{S}}^{+}$ [27]. In the case where the columns are linearly independent, thus requiring that there are at least as many isotopic measurements as there are sources, the pseudo-inverse can be computed as $\tilde{\mathbf{S}}^{+} = (\tilde{\mathbf{S}}^{\mathrm{T}}\tilde{\mathbf{S}})^{-1}\tilde{\mathbf{S}}^{\mathrm{T}}$ [28]. Therefore, when the inverse exists this computation reduces to $\tilde{\mathbf{S}}^{+} = \tilde{\mathbf{S}}^{-1}$. The distribution of the least squares solution can be expressed as a truncated normal distribution constrained to the positive half-plane, whose variance is smaller than that of the unbounded normal.

$$\tilde{\mathbf{f}} \sim \mathcal{N}_{d+1}\left(\tilde{\mathbf{S}}^{+}\tilde{X}, \eta^2(\tilde{\mathbf{S}}^{\mathrm{T}}\tilde{\mathbf{S}})^{-1}\right) \tag{2.17}$$

Furthermore, given values for the isotopic signature of the sources $S_1, \ldots, S_K$ and fractionation factor A are subject to measurement uncertainty as well as epistemic uncertainty, since their values can depend on unknown soil and environmental characteristics. It is thus natural to model these uncertainties as normally distributed random variables as well and derive the estimation uncertainty with their consideration.

$$S_k \sim \mathcal{N}_d(b_k, \xi_k^2), \quad \forall k = 1, \ldots, K$$

$$A \sim \mathcal{N}_d(c, \nu^2)$$

The distribution of the measurements X is thus a convolution of weighted Gaussian distributions, which itself is a Gaussian distribution again [29]. Inference and uncertainty quantification is then analogous to above, while using the calculated mean and variance of the measurement distribution.

$$X_j \sim \mathcal{N}\left(\sum_{k=1}^{K} f_k b_{kj} + c_j \log r,\ \sum_{k=1}^{K} f_k^2 \xi_{kj}^2 + \nu_j^2 \log^2 r + \eta^2\right), \quad \forall j = 1, \ldots, d \tag{2.18}$$

Oftentimes, the epistemic uncertainty of source isotopic signatures is modeled as a uniform distribution. It is worth to note that in the case of many similar sources a Gaussian approximation is applicable and sufficient for most applications, but it is not universally adequate. In those cases Bayesian statistics is useful to incorporate all assumptions and constraints into the model as well as to employ numerical inference methods for source contribution estimation and uncertainty estimation. This widens the framework to arbitrary mixing equations $\mu(\mathbf{f}, r)$ as well as arbitrary prior distributions for the auxiliary parameters $\mathbf{S}$ and A.

2.2. Bayesian Hierarchical Models

Bayesian models assume that the measurements as well as the parameters are associated with respective probability distributions. The concept of probability is to be understood as a measure of certainty regarding the actual value of a variable rather than the frequentist perspective that focuses on the likelihood of an outcome given repeated experiments. The Bayesian framework combines parameter estimation with uncertainty quantification and makes use of estimation methods that are applicable to a wide range of possible models.

A given model is parameterized by the vector of parameters $\theta \in \Theta$ that are equipped with a given prior probability distribution $\pi(\theta)$. In the case of isotopic source contribution and fractionation quantification the parameter vector would be $\theta = \binom{\mathbf{f}}{r}$. The statistical model then defines the likelihood of measuring a certain observation $X \in \mathbb{R}^d$ conditioned on the parameter value denoted by $p(X|\theta)$. Now, *Bayes' Theorem* can be used to derive the posterior distribution of the parameters given the observation [30].

$$\pi(\theta|X) = \frac{p(X|\theta)\pi(\theta)}{p(X)} \tag{2.19}$$

This posterior probability distribution conveys much more information than a simple maximum likelihood estimation of the parameter would. The core task of Bayesian parameter estimation is then to evaluate the integral $p(X) = \int p(X|\theta)\pi(\theta)d\theta$ or respectively to sample from this posterior distribution directly without normalization.

Bayesian hierarchical modeling adds additional layers to the prior structure, thus allowing for grouping of the measured variables X in certain dimensions [31]. In general, the choice of prior is arbitrary and numerous methods of selecting or estimating it are offered within the literature [32]–[35]. Principally, the prior structure can be utilized to encode information about the sampling populations that the measurements X originate from. Assuming that there is a set of populations warrants modeling them with potentially different prior distributions $\pi(\theta|\phi)$ indexed by the *hyperparameter* $\phi \in \Phi$. In a fully Bayesian framework this index parameter is equipped with its own probability distribution $\pi(\phi)$ which is known as a *hyperprior*.

In an empirical Bayesian framework the hyperparameter is estimated to maximize the evidence $p(X)$ beforehand, which is equivalent to the maximum a posteriori solution $\hat{\phi}$ for a flat hyperprior [36]. Alternatively, for very peaked posterior distributions, frequentist techniques such as maximum likelihood estimation can be used as well [37]. A posterior of the model parameters can then be formulated by plugging in this estimator $\pi(\theta|X) \propto p(X|\theta)\pi(\theta|\hat{\phi})$. However, most Bayesian estimation methods are reasonably efficient for a large number of parameters and thus a fully hierarchical Bayesian framework which evaluates a joint posterior $\pi(\theta, \phi|X) \propto p(X|\theta)\pi(\theta|\phi)\pi(\theta)$ is frequently the framework of choice [38].

2.2.1. Bayesian Mixing Models

For scenarios involving only isotope mixing, a Bayesian framework can be established by assuming some prior distribution $\pi(\mathbf{f})$ for the source contributions $\mathbf{f} \in \mathcal{S}_K$. The likelihood $p(X|\mathbf{f})$ is then a normal distribution centered around the result of the pure mixing equation $\mu(\mathbf{f}) = \mathbf{S}\mathbf{f}$. The matrix of source isotopic signatures $\mathbf{S} \in \mathbb{R}^{d\times K}$ is an auxiliary parameter with values taken from specific analyses of each individual source found in the literature. When uncertainties are involved, they are incorporated into a prior distribution $\pi(\mathbf{S})$ as well.

Bayesian mixing models allow for general purpose inference of mixing weights. By combining the priors for source contribution weights and source isotopic signatures, a posterior distribution can be evaluated that gives pointwise estimates of the parameters as well as an estimation of uncertainty.

$$\pi(\mathbf{f}, \mathbf{S}|X) \propto p(X|\mathbf{f}, \mathbf{S})\pi(\mathbf{f})\pi(\mathbf{S}) \tag{2.20}$$

The auxiliary parameters corresponding to source isotopic signatures $\mathbf{S}$ are latent variables and the posterior estimates their values jointly with $\mathbf{f}$. In order to get estimates of the relevant mixing weight parameters $\mathbf{f}$ only, the posterior can be marginalized by solving the integral $\pi(\mathbf{f}|X) = \int \pi(\mathbf{f}, \mathbf{S}|X)d\mathbf{S}$. Latent variables can also be integrated over when formulating the likelihood by computing $p(X|\mathbf{f}) = \int p(X|\mathbf{f}, \mathbf{S})\pi(\mathbf{S})d\mathbf{S}$. This results in a model with fewer parameters and produces the exact same marginal posterior.

$$\pi(\mathbf{f}|X) \propto p(X|\mathbf{f})\pi(\mathbf{f}) = \int p(X|\mathbf{f}, \mathbf{S})\pi(\mathbf{S})\pi(\mathbf{f})d\mathbf{S} \propto \int \pi(\mathbf{f}, \mathbf{S}|X)d\mathbf{S} \tag{2.21}$$

Many isotopic problems require solving pure mixing equations exclusively, which is why many different approaches for this setting exist. However, in the case of nitrous oxide measurements and similar situations, the endmember isotopic signature is not only governed by the mixing of different sources, but it is simultaneously removed by other processes that introduce fractionation effects. For these situations Bayesian mixing models are extended with augmented mixing equations as in Equation 2.12 using an additional parameter $r \in [0, 1]$ for the fractionation weight and equipping it with a suitable prior as well $\pi(r)$. The joint posterior can then be evaluated by modeling the fractionation factor as a random variable $A \in \mathbb{R}^d$ similar to the source isotopic signatures.

$$\pi(\mathbf{f}, r, \mathbf{S}, A|X) \propto p(X|\mathbf{f}, r, \mathbf{S}, A)\pi(\mathbf{f})\pi(r)\pi(\mathbf{S})\pi(A) \tag{2.22}$$

2.2.2. Choosing Prior Distributions

Bayesian hierarchical models are based on a representation of the data generating process that allows one to find posterior estimates of the parameters. These estimates are the consequence of a combination of prior and likelihood distributions, that must be modeled somehow. It is oftentimes clear how to construct suitable likelihood distributions since they stem from the physical modeling of the situation and are part of frequentist statistics as well. In isotopic data analysis the likelihood is given through the mixing and fractionation equation. Prior distributions on the other hand are much more flexible and are desirably chosen to reflect already known information. If this is not available so called *uninformative priors* are chosen that simply constrain the parameters to certain regions without influencing the final estimate otherwise. Claims of any distribution being uninformative are patently up to debate and ultimately the choice of the modeler. A common practice when using Rayleigh fractionation is to assign a uniform prior to the fractionation weight, although, considering its logarithmic dependence on the measurement outcomes, it is not immediately clear why this is deemed a favorable choice.

Typically, the choice of priors is separated into uninformative [35], weakly informative [39] and strongly informative priors [34]. In the presence of expert information or strong a priori results the choice of strongly informative prior should be natural and the possibility to incorporate such information is one of the major strengths of Bayesian modeling. Mixing problems for instance have notions of concentration, where priors do not presuppose that any parameter is larger than another but simply that one parameter dominates against the others [40]. In the absence of prior information the use of uninformative priors is oftentimes discouraged. This is due to the ambiguity of what it means to be uninformative and the unintended consequences and pitfalls of such assumptions [39], [41]. In certain cases such as normal-normal hierarchical models guidelines for choosing priors exist [42] and predictive checks can help in general to understand the effect a certain prior has on the estimates [33].

The *Jeffreys prior* [43] offers a way of constructing an uninformative or objective prior distribution based on the argument that it is invariant under change of coordinates and thus does not favor any particular values over others. The Jeffreys prior for a parameter θ is chosen proportionally to the square root of the *Fisher information* determinant.

$$\pi(\theta) \propto \sqrt{|\mathcal{I}(\theta)|} \tag{2.23}$$

The Fisher information is defined as the variance of the first derivative of the log-likelihood, which under certain regularity conditions can be expressed by $I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log p(X|\theta)\middle|\theta\right]$ and therefore quantifies a notion of curvature or sensitivity of the likelihood around a parameter value [44, ch. 5]. Consequently, it is modified under any transformation of the parameter $\varphi(\theta)$ to $I_\varphi(\varphi) = I_\theta(\theta)\left(\frac{d\theta}{d\varphi}\right)^2$ such that the Jeffreys prior assigns equal probability to the same volume of probability space in either parametrization $\pi(\varphi) = \pi(\theta)\left|\frac{d\theta}{d\varphi}\right|$ [45]. This property is mostly desired for parameters relating to scale [46].

2.3. Monte Carlo Methods

Bayesian estimation of source contribution and fractionation weights for instance works by constructing a prior distribution $\pi(\theta)$ for the parameters $\theta \in \Theta$ as well as a likelihood distribution $p(X|\theta)$ for the measurements $X \in \mathbb{R}^d$ and then carefully evaluating both, using Bayes' Theorem to compute a posterior distribution $\pi(\theta|X)$. Various methods exist to approximate this distribution numerically of which *Monte Carlo* methods are most prominent. A Monte Carlo method is defined as any computational algorithm that relies on random sampling to compute an output, which therefore by construction is also random [47]. They are capable of both numerical integration as well as sampling from difficult probability distributions.

Assuming that the distribution of a parameter has distribution $\theta \sim \mathcal{P}_\Theta$ with density $\pi(\theta)$, Monte Carlo integration approximates the marginal likelihood integral with the random sampling average over S identically distributed variables $\theta^{(1)}, \ldots, \theta^{(S)} \sim \mathcal{P}_\Theta$.

$$p(X) = \int_\Theta p(X|\theta)\pi(\theta)d\theta \approx \frac{1}{S}\sum_{s=1}^{S} p\left(X\middle|\theta^{(s)}\right) \tag{2.24}$$

The marginal likelihood has important statistical meaning since it can be used for model comparison and to examine the posterior numerically, although this approximation presupposes the ability to sample from $\mathcal{P}_\Theta$ instead. When this is not given, Monte Carlo methods can be used to sample from the posterior directly. Rejection sampling [48], [49] and importance sampling [50] both provide a way of adapting a sample from a suitable and simple proposal distribution to the target posterior distribution, although with varying efficiency. However, due to its wide applicability and adaptive nature, Markov Chain Monte Carlo sampling is the method of choice in the vast majority of cases.

2.3.1. Markov Chain Monte Carlo

Given the objective of Monte Carlo sampling to generate statistically efficient random sample from a challenging target distribution, it is intuitive to construct methods that use independent samples from auxiliary distributions, such as the uniform or Gaussian distribution, which offer simpler sampling procedures. *Markov Chain Monte Carlo (MCMC)* methods abandon the idea of creating a statistically independent sample in order to create a more efficient proposal distribution [51]. The result are algorithms that trade off the statistical efficiency of independent samples for

the efficiency stemming from samples having high target probability and thus large contributions to the integral in Equation 2.24, typically resulting in an overall better sample especially for high-dimensional parameter spaces.

A *Markov chain* or *Markov process* is a statistical process comprising a series of random variables $\theta^{(1)}, \theta^{(2)}, \ldots$ and is defined by the conditional probability $\mathbb{P}\left[\theta^{(s)} | \theta^{(1)}, \ldots, \theta^{(s-1)}\right]$ of each subsequent member of the chain [51, ch. 4]. To be classified as Markov process, it must satisfy the Markov property, that is to say that the probability of each observation is solely dependent on the last member of the chain and independent of all others.

$$\mathbb{P}\left[\theta^{(s)} \middle| \theta^{(1)}, \ldots, \theta^{(s-1)}\right] = \mathbb{P}\left[\theta^{(s)} \middle| \theta^{(s-1)}\right]$$

In order to construct a Markov chain that samples a specific target density $\pi(\theta)$ known only in unnormalized form $q(\theta) = Z\pi(\theta)$ for $Z > 0$, one must choose the transition probabilities $\mathbb{P}\left[\theta^{(s)} \middle| \theta^{(s-1)}\right]$ accordingly. The *Metropolis-Hastings algorithm* precisely constructs a transition probability satisfying detailed balance, which is a sufficient condition for the Markov chain to converge to the target probability distribution [30, ch. 10.3].

The algorithm then works by iteratively sampling from a proposal distribution $Q(\cdot|\theta^{(s)})$ and then evaluating an acceptance probability α that modifies the overall transition probability such that the desired distribution is sampled.

Algorithm 1: Metropolis-Hastings algorithm

Data: Initial value $\theta^{(0)}$, sample size S

for $s = 0, \ldots, S$ **do**
  Sample $\theta' \sim Q(\cdot|\theta^{(s)})$;
  Sample $u \sim \mathrm{Uni}(0, 1)$;
  Compute $\alpha = \min\left\{1, \frac{q(\theta')Q(\theta^{(s)}|\theta')}{q(\theta^{(s)})Q(\theta'|\theta^{(s)})}\right\}$;
  if $u \leq \alpha$ **then**
    $\theta^{(s+1)} \leftarrow \theta'$;
  else
    $\theta^{(s+1)} \leftarrow \theta^{(s)}$;
  end
end

Constructing proposal probabilities is not difficult in general, since it only must be ensured that all regions of the variable space Θ can be reached [30, ch. 10]. However, it is not trivial to find distributions that are also efficient, meaning they generate proposals with high acceptance rates and thus frequent transitions. Straightforward proposal distributions are the uniform distribution over the whole space Θ as well as the random-walk proposal using a Gaussian distribution centered around the previous value. In the context of sampling from Bayesian posterior distributions it also common to choose the independent prior distribution, if the posterior is expected to be similar due to the data conveying only little information.

Trivial proposal distributions can already be efficient enough depending on the problem at hand. Situations where the variable space Θ has high-dimension or where the unnormalized density $q(\theta)$ has a difficult shape are often mentioned as cases where these approaches are not sufficient anymore [52]. It is thus useful to construct more sophisticated proposal distributions. *Gibbs sampling* uses the fact the conditional distributions are perfect proposals in the sense that they will always be accepted [53, ch. 11.3]. The conditional distribution of of one single parameter θ_i in the parameter vector $\theta = [\theta_1, \ldots, \theta_n]^{\mathrm{T}} \in \Theta$ is the distribution conditioned on the parameter vector including all other parameters.

$$\pi(\theta_i|\theta_{-i}) = \pi(\theta_i|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$$

Even though the joint distribution $\pi(\theta)$ is not known, thus prompting the use of a sampler in the first place, the conditional distributions are oftentimes easy to derive analytically or to reasonably approximate with Markov blankets.

Gibbs sampling follows the same procedure as the general Metropolis-Hastings algorithm with each step corresponding to sampling only a single component θ_i' of the parameter vector and with its conditional distribution as the proposal $Q(\theta'|\theta) = \pi(\theta_i'|\theta_{-i})$. One can easily see that the acceptance step in that case becomes trivial and can be omitted, thus eliminating any issues stemming from low acceptance probabilities.

$$\frac{\pi(\theta')Q(\theta^{(s)}|\theta')}{\pi(\theta^{(s)})Q(\theta'|\theta^{(s)})} = \frac{\pi(\theta_i', \theta_{-i}^{(s)})\pi(\theta_i^{(s)}|\theta_{-i}^{(s)})}{\pi(\theta_i^{(s)}, \theta_{-i}^{(s)})\pi(\theta_i'|\theta_{-i}^{(s)})} = \frac{\pi(\theta_i'|\theta_{-i}^{(s)})\pi(\theta_{-i}^{(s)})\pi(\theta_i^{(s)}|\theta_{-i}^{(s)})}{\pi(\theta_i^{(s)}|\theta_{-i}^{(s)})\pi(\theta_{-i}^{(s)})\pi(\theta_i'|\theta_{-i}^{(s)})} = 1 \tag{2.25}$$

Increasing statistical quality of an MCMC sample by reducing autocorrelation and increasing computational efficiency by proposing better values with higher acceptance rates oftentimes presents itself as a strict trade-off. However, if not only the unnormalized density $q(\theta)$ is known, but also its derivative $\frac{d}{d\theta}q(\theta)$, this additional information can be used to generate high quality proposals with minimal autocorrelation. This is accomplished by *Hamiltonian Monte Carlo (HMC)* [54] which generates proposals by simulating a physical particle trajectory in a potential corresponding to the target log-density surface.

The *Hamiltonian* is an operator used in physics that computes the total energy of a physical system. In the case of non-interacting classical particles the energy is given by the sum of kinetic and potential energy components. The position represents the sampled parameters $\theta \in \mathbb{R}^d$, which will be accompanied by an equally sized momentum vector $\xi \in \mathbb{R}^d$. Jointly, they uniquely define the state of a system and can be used together with a fixed mass matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$ to compute the Hamiltonian.

$$H(\theta, \xi) = \frac{1}{2}\xi^{\mathrm{T}}\mathbf{M}^{-1}\xi - \log q(\theta) \tag{2.26}$$

Physical system dynamics are now described by *Hamilton's equations*. New state vectors can be generated and proposed for the sampling algorithm by following these parameter dynamics $\theta(t)$ and $\xi(t)$ for a suitable amount of time.

$$\frac{d\theta}{dt} = \frac{\partial H}{\partial \xi}, \quad \frac{d\xi}{dt} = -\frac{\partial H}{\partial \theta} \tag{2.27}$$

Hamilton's equations can be solved numerically by discretizing in time and employing for instance the leap-frog algorithm to integrate up until a final time T [53, ch. 11.5]. The previous sample will be set as the initial condition $\theta(0) = \theta^{(s)}$ and the position at the end of the path $\theta' = \theta(T)$ is proposed as new sample and subjected to the acceptance step. Improvements to the sampling strategy exist, such as the *No U-Turn Sampler* [55], [56], which selects the stopping time T adaptively. Since the path of a particle is restricted inside the potential by its initial energy, the initial momentum is randomized and typically sampled from a Gaussian distribution $\xi(0) \sim \mathcal{N}_d(0, \mathbf{M})$. Both the sampling position θ as well as the momentum ξ are included in the modified acceptance probability, thereby ensuring that the accepted parameters follow the target distribution and the random particle states in each step follow a Boltzmann distribution [57].

$$\alpha = \min\left\{1, \frac{\exp(-H(\theta(T), \xi(T))}{\exp(-H(\theta(0), \xi(0))}\right\} \tag{2.28}$$

2.3.2. Evaluating Sample Quality

Markov chain Monte Carlo simulation is used to create a correlated sample $\theta^{(1)}, \ldots, \theta^{(S)}$ which is then used to estimate posterior distribution properties such as its mean for pointwise estimation and quantiles for uncertainty estimation. However, statistical samplers have many failure modes in which the produced sample does not represent the posterior well. Theoretical guarantees of the Metropolis-Hastings procedure exist only for an unbounded number of samples [58]. For this reason it is important to evaluate the sample quality after a Monte Carlo simulation is done.

For the estimation of distribution properties independent samples are typically preferred. Markov chain Monte Carlo methods inherently produce subsequent samples on the basis of previous values and thus create dependence. In the worst case S posterior samples are produced that are heavily correlated and thus do not give accurate estimates of its distribution properties. The autocovariance of a statistical process such as a stationary Markov chain is defined as the covariance between elements having lag k between them.

$$\gamma_k := \mathrm{Cov}\left[\theta^{(s)}, \theta^{(s+k)}\right] \tag{2.29}$$

Distribution estimators derived from independent samples have estimation errors proportional to $\frac{1}{\sqrt{N}}$ for N observations [59]. This fact is a consequence of the linearity of variance which holds for independent random variables but not for correlated ones. It can be shown that the total variance of a correlated sample estimate is given by the sum of autocovariance coefficients $\sum_{k=-\infty}^{\infty} \gamma_k$ which can be evaluated using autoregressive processes or Fourier transforms [60]. A notion of *effective sample size* is then given by the actual number of observations S scaled by the ratio of the correlated sample variance with the empirical variance estimator $\hat{\sigma}^2$.

$$N_{eff} = S\frac{\hat{\sigma}^2}{\sum_{k=-\infty}^{\infty} \gamma_k} \tag{2.30}$$

The effective sample size N_{eff} now corresponds to the estimation error using the correlated sample in the analogous way, having errors proportional to $\frac{1}{\sqrt{N_{eff}}}$ [61].

Since Markov chains can fail to sample the full posterior in the presence of stiffness or multimodality, it is recommended to run multiple chains and assess their similarity to determine if the generated sample is well mixed and stationary. Let the sample be replicated in M chains with values $\theta^{(m,1)}, \ldots, \theta^{(m,S)}$ for $m = 1, \ldots, M$. Each chain can compute an individual mean estimate $\hat{\mu}_m = \sum_{s=1}^{S} \theta^{(m,s)}$ and variance estimate $\hat{\sigma}_m^2 = \frac{1}{S-1} \sum_{s=1}^{S} (\theta^{(m,s)} - \hat{\mu}_m)^2$. The within-chain variance is then given by the average $W = \frac{1}{M} \sum_{m=1}^{M} \hat{\sigma}_m^2$ and the between-chain variance is given by $B = \frac{S}{M-1} \sum_{m=1}^{M} (\hat{\mu}_m - \bar{\mu})^2$ for $\bar{\mu} = \frac{1}{M} \sum_{m=1}^{M} \hat{\mu}_m$, which can be used to compute the *Gelman-Rubin statistic* as the ratio between weighted combined variances to the average within-chain variance [62].

$$\widehat{R} = \sqrt{\frac{\frac{S-1}{S} W + \frac{1}{S} B}{W}} \tag{2.31}$$

If the chains are slowly mixing, the overall variance of the combined sample will exceed the average within-chain variance and a high $\widehat{R}$ indicates sampling issues. The Gelman-Rubin statistic is considered acceptable and the chains are declared as stationary and converged if $\widehat{R} < 1.2$ meaning that the variance increase over within-chain variances is smaller than 20% [63]. Nevertheless, further failure modes can occur that this criterion cannot detect and several variations for measuring mixing quality and stationarity exist [64], [65].

2.3.3. Statistical Sampling Software

Monte Carlo simulation is almost universally used to implement Bayesian models and thus there are numerous libraries in various programming languages assisting with this task. The implementation of statistical sampling, which is largely independent of the specific model being evaluated, allows sampling libraries to interface with high-level model descriptions and produce posterior samples using a general purpose strategy. Having robust software already in place, that can deal with the complexities of implementing efficient Monte Carlo sampling, especially for high-dimensional parameter spaces, is incredibly valuable for the purpose of developing and implementing novel tools designed to analyze isotopic data.

Stan is a library implementing a high-performance Hamiltonian Monte Carlo sampler written in `C++` for Bayesian modeling [66]. It contains many utility functions for complex operations and can be used for variational Bayesian inference and penalized maximum likelihood estimation as well. Stan additionally has features helping with the assessment of sampler health and data quality. Most notably it is used in packages implementing Bayesian linear regression techniques.

Just Another Gibbs Sampler (JAGS) is a Gibbs sampling implementation in `C++` based on the BUGS language [67]. It is fairly lightweight and does not possess data post-processing features itself, but these are usually available in statistical programming languages used in conjunction. Many high level Bayesian analysis packages use JAGS as a sampler in the background and it can tackle a wider range of problems such as modeling integer parameters.

Interfaces to the `R` programming language exist for both samplers, making tool development and distribution fairly straightforward and allowing users to directly integrate Monte Carlo samplers in their data analysis workflow [68]–[70].

2.4. Smoothing

The numerical and Bayesian techniques to estimate source contribution and fractionation weights based on isotopic measurements rely on accuracy of the input data in order to derive the estimates. However, it can be the case that the measurement error becomes overwhelming making precise inference impossible. Time series data $X(\tau) \in \mathbb{R}$ can be preprocessed to eliminate some of this error variation by assuming regularity and introducing the series of smooth values $W(\tau) \approx X(\tau)$ that balances regularity and accuracy in hopes of eliminating noise.

For sampling points $\tau_1, \ldots, \tau_N$ of the measurement time series $X_i := X(\tau_i)$ the smooth approximation can rely on different points in time $\tilde{\tau}_1, \ldots, \tilde{\tau}_M$ to compute the estimation, which can sometimes be advantageous for the construction of smooth estimates $W_j = W(\tilde{\tau}_j)$. Generally, using less evaluation points $M < N$ reduces variation, because more raw points are aggregated.

2.4.1. Linear Smoothing Operators

Linear smoothing operators assume that the estimated values $W_j, j = 1, \ldots, M$ are a linear combination of the original measurements $X_i, i = 1, \ldots, N$. Thus, arranging both series in vector form $W := [W_1 \cdots W_M]^{\mathrm{T}} \in \mathbb{R}^M$ and $X := [X_1 \cdots X_N]^{\mathrm{T}} \in \mathbb{R}^N$, linear smoothing operators are matrices $\mathbf{G} \in \mathbb{R}^{M \times N}$ such that the estimates can be computed as $W = \mathbf{G}X$.

Using the idea of a local average, kernel smoothing operators can be derived. Stemming from kernel density estimation the *Nadaraya-Watson estimator* is given by the weighted average of the raw data points, where the weights are computed using a symmetric kernel function $\kappa(\cdot)$ [71].

$$W(\tau) = \frac{\sum_{i=1}^{N} \kappa(\tau_i - \tau) X(\tau_i)}{\sum_{i=1}^{N} \kappa(\tau_i - \tau)} \tag{2.32}$$

In order to derive the matrix formulation, the linear operator can be constructed with the elements $G_{mi} = \frac{\kappa(\tau_i - \tilde{\tau}_m)}{\sum_{j=1}^{N} \kappa(\tau_j - \tilde{\tau}_m)}$ for a fixed set of evaluation points of the smoother $\tilde{\tau}_1, \ldots, \tilde{\tau}_M$.

$$W_m = \sum_{i=1}^{N} \frac{\kappa(\tau_i - \tilde{\tau}_m)}{\sum_{j=1}^{N} \kappa(\tau_j - \tilde{\tau}_m)} X_i = \sum_{i=1}^{N} G_{mi} X_i \quad \forall m = 1, \ldots, M \tag{2.33}$$

This kernel smoothing operator encompasses for instance moving averages in the case of equally spaced time series. An extension to this construction are local linear and local polynomial regression estimators such as *LOESS* [72]. They estimate derivatives alongside the smooth time series and can take uncertainties into account.

A different approach to the problem of finding a smooth representation $W(\tau)$ of a time series $X(\tau)$ can be taken by assuming that the function is differentiable up to a certain order and thus it can be expressed as a spline polynomial [73, ch. 1.4]. Spanning the space of splines using a suitable basis $s_1(\cdot), \ldots, s_M(\cdot)$ with restricted degrees of freedom M, the smooth function is given as a linear combination.

$$W(\tau) = \sum_{j=1}^{M} b_j s_j(\tau) \tag{2.34}$$

Evaluations of the basis functions at the measurement times $\tau_1, \ldots, \tau_N$ can be arranged in a design matrix $H_{ij} := s_j(\tau_i)$. The vector of basis function coefficients $b = [b_1 \cdots b_M]^\mathrm{T} \in \mathbb{R}^M$ can now be chosen as the least squares estimate in order to minimize the approximation error of the spline polynomial. This is exactly the least squares regression solution for the design matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ [74, ch. 10.2].

$$\hat{b} = \underset{b \in \mathbb{R}^M}{\operatorname{argmin}} \|X - \mathbf{H}b\|^2 = (\mathbf{H}^\mathrm{T}\mathbf{H})^{-1}\mathbf{H}^\mathrm{T}X \tag{2.35}$$

Now the smoothing operator can be expressed as the projection to the space of spline function evaluations $W = \mathbf{G}X = \mathbf{H}(\mathbf{H}^\mathrm{T}\mathbf{H})^{-1}\mathbf{H}^\mathrm{T}X$.

It is important to make the distinction between spline regression and smoothing splines, which are arguably more widely used for this particular purpose. They use an equal number of basis functions as there are measurement points $M = N$ but enforce smoothness via a penalty on the second derivatives, which translates to a generalized least squares estimate [75].

2.4.2. Gaussian Processes

Gaussian process regression is a smoothing technique using the assumption that the time series $X(\tau)$ is jointly normally distributed for all discrete subsets of points in time, and thus inference of the distribution means gives an estimate with reduced noise. In a Bayesian interpretation of this setting, a Gaussian process prior distribution is applied on the smooth means $W(\tau)$.

The Gaussian process prior distribution is defined over a vector of function evaluations $W = [W_1 \cdots W_N]^\mathrm{T} = [W(\tau_1) \cdots W(\tau_N)]^\mathrm{T} \in \mathbb{R}^N$. It is given by a constant mean $\mu \in \mathbb{R}$ and a covariance kernel $\mathbf{G}_{ij} = \sigma^2 \kappa_\rho(\tau_i, \tau_j)$ with scale parameter $\sigma \in \mathbb{R}_+$ and a symmetric kernel function $\kappa_\rho(\cdot, \cdot)$. Since it is usual to work with normalized data the prior is typically set to zero mean $\mu = 0$ and unit variance $\sigma^2 = 1$. Alternatively, the Gaussian process can be shifted and scaled to the correct shape by plugging in the empirical mean $\hat{\mu}_X$ and variance $\hat{\sigma}_X^2$ of the data to be fitted for an equivalent result. Examples of kernel functions include the radial basis function kernel, the exponential kernel and the Matérn kernel, which are all parameterized by a correlation length $\rho \in \mathbb{R}_+$ [76], [77].

$$\kappa_\rho(\tau_i, \tau_j) = \exp\left(-\frac{(\tau_i - \tau_j)^2}{2\rho^2}\right) \tag{2.36}$$

$$\kappa_\rho(\tau_i, \tau_j) = \exp\left(-\frac{|\tau_i - \tau_j|}{\rho}\right) \tag{2.37}$$

$$\kappa_\rho(\tau_i, \tau_j) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}|\tau_i - \tau_j|}{\rho}\right)^\nu K_\nu \left(\frac{\sqrt{2\nu}|\tau_i - \tau_j|}{\rho}\right) \tag{2.38}$$

Here, $\Gamma(\cdot)$ denotes the gamma function and $K_\nu(\cdot)$ denotes the modified Bessel function of the second kind. Different kernel functions can be used to codify assumptions about the smoothed solution into the distribution. Furthermore, different kernels can be combined to express even more sophisticated relationships.

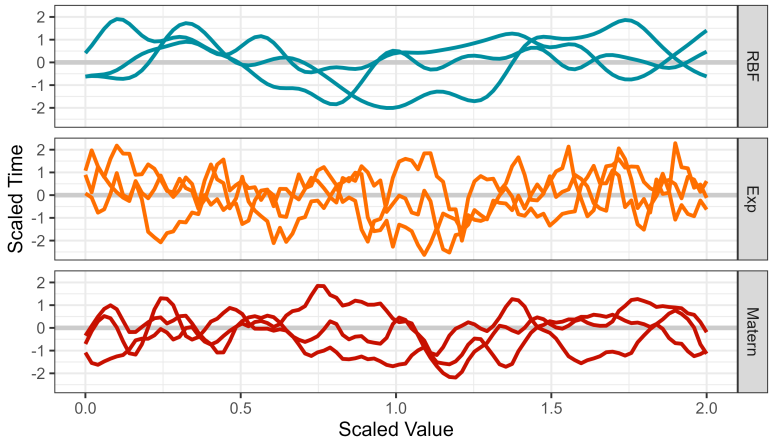

Figure 2.1.: Gaussian process prior for $W(\tau)$ sampled three times independently for radial basis function kernel, exponential kernel and Matérn kernel of order $\nu = 2$ all having correlation length $\rho = 0.1$.

The main parameter that governs the properties of the kernel covariance function is the correlation length ρ. It can be set beforehand to a suitable value, reflecting the rate of change that is expected from the smoothed time series, or it can be equipped with a hyperprior distribution and estimated in conjunction with the smooth time series values W. An empirical Bayesian estimate can then be used for efficiency or a fully hierarchical model can be formulated to estimate a joint posterior. Different approaches exist in terms of defining suitable priors. A common choice is the inverse Gamma distribution, since it has an appropriate tail and is zero avoiding.

The Bayesian model can now be formulated using the Gaussian process prior and independent Gaussian distributions with variance $\eta \in \mathbb{R}_+$ as likelihood. Because vectors with components belonging to different points in time will later be reordered to row-vectors the Gaussian process distribution shall be defined as a row-vector distribution and therefore must be transposed here.

$$\begin{aligned} W &\sim \mathcal{GP}^{\mathrm{T}}(\mu, \sigma, \mathbf{G}) = \mathcal{N}_N(\mu, \sigma^2\mathbf{G}) \\ X &\sim \mathcal{N}_N(W, \eta^2\mathbb{1}) = \mathcal{N}_N(\mu, \eta^2\mathbb{1} + \sigma^2\mathbf{G}) \end{aligned} \tag{2.39}$$

For measurement series it is often the case that multiple values exist at the same time and thus a single time step is represented by a d-dimensional measurement vector $X_t \in \mathbb{R}^d$ and consequently smooth means $W_t \in \mathbb{R}^d$. In this situation the values are arranged into a measurement matrix $\mathbf{X} = [X_1 \cdots X_N] \in \mathbb{R}^{d\times N}$. Expressing the mean, scale and noise of the Gaussian process as vectors $\mu, \sigma, \eta \in \mathbb{R}^d$, the row-vectors $\mathbf{W}_{j:}$ have independent Gaussian process distributions.

$$\begin{aligned} \mathbf{W}_{j:} &\sim \mathcal{GP}(\mu_j, \sigma_j, \mathbf{G}) \quad \forall j = 1, \ldots, d \\ \mathbf{X}_{ji} &\sim \mathcal{N}(\mathbf{W}_{ji}, \eta_j^2) \quad \forall j = 1, \ldots, d \quad \forall i = 1, \ldots, N \end{aligned} \tag{2.40}$$

The prior can then use the notation $\mathbf{W} \sim \mathcal{GP}_d(\mu, \sigma, \mathbf{G})$ denoting the matrix distribution with row-wise independent Gaussian processes.

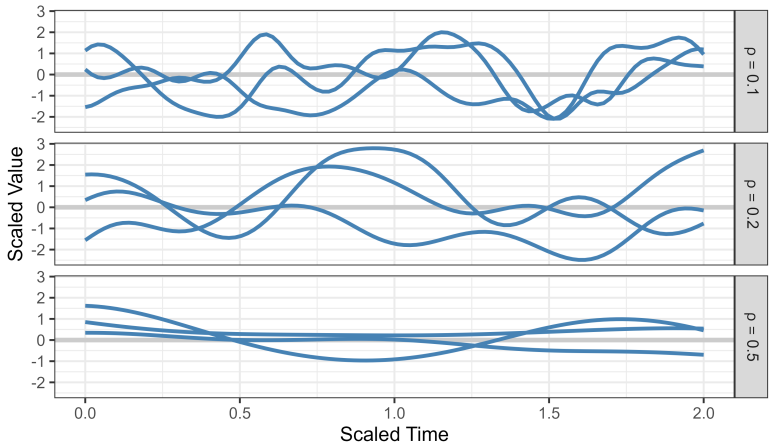

Figure 2.2.: Gaussian process prior for $W(\tau)$ sampled three times independently with radial basis function kernel for different values of correlation length ρ.

2.4.3. Multiple Dimensions

It is conceivable that isotopic measurements are taken with more than just one continuous covariate corresponding to time, for instance in the case of spatio-temporal measurements over a given area or in different depths $u_1, \ldots, u_M$. Natural extensions to Gaussian processes can easily be constructed by applying kernel functions in multiple dimensions and bases for spline regression can for instance be expanded to tensor product splines [78].

However, in order to take time series information into account it has to be implicitly assumed that the system is non-stationary, even though the exact dynamics are usually not considered and smooth estimators use notions of correlation length that are not necessarily physical in nature. Considering concentration measurements $X(u, \tau)$ over space u and time τ, their physical dynamics is given by Fick's second law for some diffusion constant $D \in \mathbb{R}$ [79].

$$\frac{\partial X}{\partial \tau} = D \frac{\partial^2 X}{\partial u^2} \tag{2.41}$$

Now accounting for source contributions $\mathbf{f}(\tau)$ and fractionation $r(\tau)$ as functions over time the production rate in terms of concentrations is summarized in the mixing and fractionation equation $\mu\big(\mathbf{f}(\tau), r(\tau)\big)$. It can be added as contribution to the change in the isotopic measurement value thus yielding an inhomogeneous diffusion equation.

$$\frac{\partial X}{\partial \tau} = D \frac{\partial^2 X}{\partial u^2} + \mu\big(\mathbf{f}(\tau), r(\tau)\big) \tag{2.42}$$

This partial differential equation can be solved for source contributions and fractionation if suitable boundary conditions are provided, and even possible analytical solutions closely related to kernel smoothers could be derived using Green's functions [80]. Possible extensions include the addition of stochastic terms to model uncertainty and extension of source contributions and fractionation to functions over space and time $\mathbf{f}(u, \tau)$ and $r(u, \tau)$.

Approaches using these assumptions are fundamentally different from the other models considered in this thesis since they rely on a differentiation between concentration and flux measurements. It is unclear whether this approach can be efficiently paired with Bayesian statistics for uncertainty quantification, but approaches using diffusion solvers have been presented for instance by Decock, Lee, Barthel, *et al.* [81].

2.5. Compositional Modeling

Solutions of the d-dimensional pure mixing equation $X = \sum_{j=1}^{K} f_j S_j \in \mathbb{R}^d$ are vectors of source contributions $\mathbf{f} = [f_1 \ \cdots \ f_K]^\mathrm{T}$ that additionally fulfill the sum condition $\sum_{j=1}^{K} f_j = 1$. For this reason, only $K-1$ degrees of freedom have to be determined and the system of equations is well defined for $d = K-1$ measurements. The domain of $\mathbf{f}$ is the *K-simplex*, which is defined as the set of K-vectors satisfying the above sum constraint as well as a non-negativity constraint.

$$\mathbf{f} \in \mathcal{S}_K := \left\{ \mathbf{f} \in \mathbb{R}^K \,\middle|\, \sum_{j=1}^{K} f_j = 1,\ f_j \geq 0 \ \ \forall j = 1, \ldots, K \right\}$$

Inference on variables on the simplex such as $\mathbf{f} \in \mathcal{S}_K$ can be done using compositional modeling techniques. Instead of relying on methods that inherently incorporate the sum constraint into their solution, a common technique is to find bijective functions from $\mathcal{S}_K$ to the Euclidean vector space $\mathbb{R}^{K-1}$, which has the appropriate degrees of freedom and thus allows for consideration of the problem in terms of unconstrained variables. In order to find mappings of the form $\mathcal{S}_K \to \mathbb{R}^{K-1}$, appropriate operations must be defined on the simplex. This can be accomplished by equipping it with the so called *Aitchison geometry* characterized by definitions of vector addition and scalar multiplication in order to form a linear space [82].

$$\mathbf{f} \oplus \mathbf{f}' := \left[\frac{f_1 f_1'}{\sum_{j=1}^{K} f_j f_j'} \ \cdots \ \frac{f_K f_K'}{\sum_{j=1}^{K} f_j f_j'} \right]^\mathrm{T} \quad \forall \mathbf{f}, \mathbf{f}' \in \mathcal{S}_K$$

$$\alpha \odot \mathbf{f} := \left[\frac{f_1^\alpha}{\sum_{j=1}^{K} f_j^\alpha} \ \cdots \ \frac{f_K^\alpha}{\sum_{j=1}^{K} f_j^\alpha} \right]^\mathrm{T} \quad \forall \alpha \in \mathbb{R} \ \ \forall \mathbf{f} \in \mathcal{S}_K$$

There are now multiple ways to define isomorphisms from the Aitchison simplex to the real space $\mathbb{R}^{K-1}$, one of which is the additive log ratio transform, that separates the last component f_K to form ratios and thus a vector of the correct dimensionality [83].

$$\mathrm{alr}(\mathbf{f}) := \left[\log \frac{f_1}{f_K} \ \cdots \ \log \frac{f_{K-1}}{f_K} \right]^\mathrm{T} \quad \forall \mathbf{f} \in \mathcal{S}_K \tag{2.43}$$

However, the components of the vector $\mathbf{f}$ are not exchangeable, since f_K is arbitrarily chosen and serves a different function than the other components. This could lead to discrepancies in estimation when component definitions are swapped, which is rarely desired. Furthermore, the

transformation is not isometric and can distort distances of mapped vectors. In order to use all components exchangeably instead of using one particular component as reference, the geometric mean of the vector can be used to compute the ratios instead.

$$g(\mathbf{f}) := \left(\prod_{i=1}^{K} f_i\right)^{\frac{1}{K}}$$

The transformation constructed this way is known as the centered log ratio [83]. Since now all components of the simplex vector are used, the mapped vector has K components as well and does not represent the correct number of dimensions.

$$\operatorname{clr}(\mathbf{f}) := \left[\log \frac{f_1}{g(\mathbf{f})} \quad \cdots \quad \log \frac{f_K}{g(\mathbf{f})}\right]^{\mathrm{T}} \quad \forall \mathbf{f} \in \mathcal{S}_K \tag{2.44}$$

Although this transformation is not bijective since it maps onto the K-dimensional Euclidean vector space, it provides a useful relation for the addition $\operatorname{clr}(\mathbf{f} \oplus \mathbf{f}') = \operatorname{clr}(\mathbf{f}) + \operatorname{clr}(\mathbf{f}')$ and multiplication $\operatorname{clr}(\alpha \otimes \mathbf{f}) = \alpha \operatorname{clr}(\mathbf{f})$. Furthermore, the image of the centered log ratio transform in $\mathbb{R}^K$ is the linear subspace $\mathcal{U}_K := \left\{\mathbf{u} \in \mathbb{R}^K : \sum_{j=1}^{K} u_j = 0\right\}$, on which it is bijective and its inverse generalized to the entire Euclidean space $\mathbb{R}^K$ is known as the softmax function. It can then be used to extend the standard inner product in the Euclidean space to the K-simplex and define the Aitchison inner product.

$$\langle \mathbf{f}, \mathbf{f}' \rangle_a := \langle \operatorname{clr}(\mathbf{f}), \operatorname{clr}(\mathbf{f}') \rangle = \frac{1}{K} \sum_{i=1}^{K} \log \frac{f_i}{g(\mathbf{f})} \log \frac{f_i'}{g(\mathbf{f}')}$$

The centered log ratio transform is in fact already an isometry with respect to the norm on the simplex induced by the Aitchison inner product $\|\mathbf{f}\|^2 = \langle \mathbf{f}, \mathbf{f} \rangle$, albeit between the simplex $\mathcal{S}_K$ and the linear subspace $\mathcal{U}_K$. An isometric transformation to $\mathbb{R}^{K-1}$ can be constructed by composition with a mapping from $\mathcal{U}_K \to \mathbb{R}^{K-1}$ using a basis expansion with respect to an orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_{K-1} \in \mathcal{U}_K$. It is not obvious how to choose a basis system, but simple standard bases can be derived from Euclidean orthonormal bases [84] or they can be specifically adapted to the data by principal balance methods [85], [86]. In fact, any orthonormal basis in the linear subspace $\mathcal{U}_K$ can expressed as the transformation from an orthonormal basis on the simplex $\mathbf{e}_1, \ldots, \mathbf{e}_{K-1} \in \mathcal{S}_K$ by the centered log ratio $\operatorname{clr}(\mathbf{e}_i) = \mathbf{u}_i$. Now any vector $\mathbf{f} \in \mathcal{S}_K$ mapped to the linear subspace $\mathcal{U}_K$ via the centered log ratio can be expanded with coefficients $\langle \operatorname{clr}(\mathbf{f}), \mathbf{u}_i \rangle$. By this construction it follows immediately that the composition of clr and the basis change with the matrix $\mathbf{U} := [\mathbf{u}_1 \cdots \mathbf{u}_{K-1}]^{\mathrm{T}} \in \mathbb{R}^{K-1 \times K}$ is characterized by coefficients $\langle \operatorname{clr}(\mathbf{f}), \mathbf{u}_i \rangle = \langle \mathbf{f}, \mathbf{e}_i \rangle_a$, which defines the isometric log ratio transform [83].

$$\operatorname{ilr}(\mathbf{f}) := \mathbf{U} \operatorname{clr}(\mathbf{f}) = [\langle \mathbf{f}, \mathbf{e}_1 \rangle_a \quad \cdots \quad \langle \mathbf{f}, \mathbf{e}_{K-1} \rangle_a] \tag{2.45}$$

Thus given any Euclidean vector $\mathbf{x} \in \mathbb{R}^{K-1}$, the inverse isometric log ratio transform is given in terms for the centered log ratio $\operatorname{ilr}^{-1}(\mathbf{x}) = \operatorname{clr}^{-1}(\mathbf{U}^{\mathrm{T}}\mathbf{x})$.

2.6. Related Research

2.6.1. General Stable Isotope Mixing Models

General stable isotope mixing models (SIMMs) were introduced by Parnell, Phillips, Bearhop, *et al.* [21] for the purpose of generalizing animal diet partitioning to cases where isotopic compositions are not observed directly. The model works with N data points of d isotopic measurements $X_i \in \mathbb{R}^d$ for $i = 1, \ldots, N$, where each data point is associated with M auxiliary measurements $Z_1, \ldots, Z_N \in \mathbb{R}^d$ that are then used to predict the source contributions. K sources are modeled with auxiliary parameters $\mathbf{S}_1, \ldots, \mathbf{S}_N \in \mathbb{R}^{d \times K}$ and fractionation corrected versions $\mathbf{C}_1, \ldots, \mathbf{C}_N \in \mathbb{R}^{d \times K}$ that have different values for each data point and are given normal distribution priors based on previously available information. Source contributions are then estimated via generalized linear model using the ilr transform as link function with parameters $\boldsymbol{\beta} \in \mathbb{R}^M$. The implementation of this model is available in the `R` package `simmr` [87].

$$\begin{aligned} \mathbf{S}_i^{\mathrm{T}} &\sim \mathcal{N}_K(\mu_s, \Sigma_s) \\ \mathbf{C}_i^{\mathrm{T}} &\sim \mathcal{N}_K(\mu_c, \Sigma_c) \\ \operatorname{ilr}(\mathbf{f}_i) &\sim \mathcal{N}_{K-1}(\boldsymbol{\beta}^{\mathrm{T}} Z_i, \boldsymbol{\kappa}) \\ X_i &\sim \mathcal{N}_d\left((\mathbf{S}_i + \mathbf{C}_i)\mathbf{f}_i, \eta^2\right) \end{aligned} \tag{2.46}$$

2.6.2. Stable Isotope Analysis in R

The stable isotope analysis in R (SIAR) framework provided the ability to do Bayesian data analysis for different isotopic problems. MixSIAR is an extension introduced by Stock, Jackson, Ward, *et al.* [20] that aims to unify modeling approaches by having flexible error structures including fixed and random effects that cover typical study designs in ecology. A total of N different measurements $X_i \in \mathbb{R}^d$ for $i = 1, \ldots, N$ are analyzed with regards to additional information $Z_1, \ldots, Z_N \in \mathbb{R}^M$ and $Y_1, \ldots, Y_N \in \mathbb{R}^L$ corresponding to fixed effects parameters $\boldsymbol{\beta} \in \mathbb{R}^M$ and random effects parameters $\mathbf{u} \in \mathbb{R}^L$. Source contribution weights are thus estimated with a generalized linear mixed-effects model using the ilr link function, Gaussian priors for fixed effects and uniform priors for random effect standard deviation. The implementation as `R` package is available as `MixSIAR` [88].

$$\begin{aligned} \boldsymbol{\sigma} &\sim \mathrm{Uni}_{K-1}[0, s] \\ \mathbf{u} &\sim \mathcal{N}_{K-1}(0, \boldsymbol{\sigma}^2) \\ \boldsymbol{\beta} &\sim \mathcal{N}_{K-1}(0, \boldsymbol{\nu}^2) \\ \operatorname{ilr}(\mathbf{f}_i) &= \boldsymbol{\beta}^{\mathrm{T}} Z_i + \mathbf{u}^{\mathrm{T}} Y_i \\ X_i &\sim \mathcal{N}_d\left(\mathbf{S}\mathbf{f}_i, \eta^2\right) \end{aligned} \tag{2.47}$$

2.6.3. Fractionation and Mixing Evaluation

The fractionation and mixing evaluation (FRAME) model [18] is a Bayesian framework for the simultaneous evaluation of mixing contributions and fractionation progress. It uses a matrix of source contributions $\mathbf{S} = [S_1 \cdots S_K] \in \mathbb{R}^{d \times K}$ to analyze d isotopic measurements $X \in \mathbb{R}^d$ while taking fractionation factors $\mathbf{A} = [A_1 \dots A_L] \in \mathbb{R}^{d \times L}$ into account. Inference is done jointly on mixing weights $\mathbf{f}$ and fractionation weights $\mathbf{r}$. Notable is that the tool can work with arbitrary mixing and fractionation equations $\mu(\mathbf{f}, \mathbf{r})$ to cover a wide range of scenarios even when effects on measurements cannot be determined by generalized linear models. The implementation is available as a user interface tool written in Python [19].

$$\begin{aligned} S_j \sim \mathrm{Uni}(b_j, \Delta_j), \;\; A_\ell \sim \mathcal{N}_d(c_\ell, \nu_\ell^2) \\ \mathbf{f} \sim \mathrm{Dir}(1), \;\; \mathbf{r} \sim \mathrm{Uni}(0, 1) \\ X \sim \mathcal{N}_d\big(\mu(\mathbf{f}, \mathbf{r}), \eta^2\big) \end{aligned} \tag{2.48}$$

Chapter 3

Methodology

3.1. Stationary Inference

Inference for the source contributions $\mathbf{f} \in \mathcal{S}_K$ and fractionation weight $r \in [0, 1]$ for one single measurement independent of time can be accomplished by the original FRAME model [18]. It constructs a prior and likelihood structure where the isotopic measurements $X \in \mathbb{R}^d$ are independently normally distributed with variance vector $\eta^2 \in \mathbb{R}_+^d$ around a mean given by an arbitrary mixing equation $\mu(\mathbf{f}, r)$. The source contributions $\mathbf{f}$ are then equipped with a flat Dirichlet prior and the fractionation weight with a uniform prior.

$$\begin{aligned} \mathbf{f} \sim \mathrm{Dir}(1), \quad r \sim \mathrm{Uni}(0, 1) \\ X|\mathbf{f}, r \sim \mathcal{N}_d\big(\mu(\mathbf{f}, r), \eta^2\big) \end{aligned} \tag{3.1}$$

The auxiliary parameters for the mixing equation $\mathbf{S}$ and A are understood to be random variables as well with predetermined fixed priors that are omitted from the model description above. Choosing those priors is ultimately up to the origin of the data and thus not subject to the further engineering of inference models in the sections below. The likelihood of X is understood to be implicitly conditional on these auxiliary parameters. This means that a joint posterior $\pi(\mathbf{f}, r, \mathbf{S}, A|X)$ is fit by the model and the reported posterior $\pi(\mathbf{f}, r|X)$ is simply its marginalization.

$$\pi(\mathbf{f}, r|X) = \int \pi(\mathbf{f}, r, \mathbf{S}, A|X) d\mathbf{S} dA \propto \int p(X|\mathbf{f}, r, \mathbf{S}, A)\pi(\mathbf{f}, r)\pi(\mathbf{S}, A) d\mathbf{S} dA$$

Different likelihood formulations exist, where the auxiliary mixing parameters are already integrated. This can be done by solving the integral $p(X|\mathbf{f}, r) = \int p(X|\mathbf{f}, r, \mathbf{S}, A)\pi(\mathbf{S}, A) d\mathbf{S} dA$ beforehand, which is analytically possible for instance if the auxiliary priors are Gaussian. Using the proper marginal likelihood can be an improvement due to the reduction in parameters that need to be fit by the sampler. However, the number of auxiliary parameters stays constant when the model is extended to multiple points in time and thus this improvement becomes less relevant for time series models. It is worth to note that the inferred posterior is theoretically the same with the only difference being efficiency of computation and the potential interpretation of posterior predictive distributions.

$$\pi(\mathbf{f}, r|X) \propto p(X|\mathbf{f}, r)\pi(\mathbf{f}, r) = \int p(X|\mathbf{f}, r, \mathbf{S}, A)\pi(\mathbf{f}, r)\pi(\mathbf{S}, A) d\mathbf{S} dA$$

3.1.1. Priors for Stationary Parameters

The FRAME model can be extended by taking different choices of prior distributions for the parameters of interest, namely the source contributions $\mathbf{f}$ and the fractionation weight r, into account. The Jeffreys prior for source contributions is constructed by computing the Fisher information matrix and choosing the probability distribution proportional to the square root of its determinant. For the source contributions $\mathbf{f} = \begin{pmatrix} 1-f \\ f \end{pmatrix} \in \mathcal{S}_2$ of two sources $S_1, S_2 \in \mathbb{R}$ the computation can be done by omitting the influence of fractionation.

$$\mathcal{I}_f(f) \propto -\mathbb{E}\left[\frac{d^2 f}{df^2}\frac{(X - S_1(1-f) - S_2 f)^2}{\eta^2}\middle| f\right] = \frac{2(S_1 - S_2)^2}{\eta^2} \propto 1 \tag{3.2}$$

Therefore the objective Jeffreys prior is uniform over the domain of $\mathbf{f}$. By symmetry this extends to multiple source contributions $\mathbf{f} \in \mathcal{S}_K$, which is equivalent to the flat Dirichlet distribution already used in the original FRAME model.

Taking now Rayleigh fractionation with weight $r \in [0, 1]$ independently of the mixing weights $\mathbf{f}$ into account, the Jeffreys prior can be computed relative to the pure mixing solution $M = X - S_1(1-f) - S_2 f \in \mathbb{R}$ with fractionation factor $A \in \mathbb{R}$.

$$\mathcal{I}_r(r) \propto -\mathbb{E}\left[\frac{d^2 r}{dr^2}\frac{(M - A\log r)^2}{\eta^2}\middle| r\right] = \frac{2A^2}{\eta^2 r^2} \propto \frac{1}{r^2} \tag{3.3}$$

Therefore the objective prior for the Rayleigh fractionation weight is given by $\pi(r) \propto \frac{1}{r}$ for $r \in [0, 1]$, which is also known as the logarithmic prior and since it cannot be normalized it is an improper prior. Additionally, even though this prior can be considered uninformative for r individually according to the Jeffreys criterion, a joint prior could lead to different results, although priors are typically chosen as independent distributions.

In the case of Rayleigh fractionation, however, it might be more reasonable to use a different distribution that is somewhere in between the uniform and logarithmic prior and incorporates the bounds to the interval $[0, 1]$ as well, which is a similar idea to prior averaging [89]. The beta distribution offers a functional form that is similar to the Jeffreys prior, but can be normalized. Parameterizing the distribution with a restricted concentration parameter $\alpha \in [0, 1]$, the form $\mathrm{Beta}(\alpha, 1)$ is the uniform distribution for $\alpha = 1$ and converges to the Jeffreys prior for $\alpha \to 0$, thus expressing a generalized approach.

3.2. Time Series Inference

In order to incorporate time series information in the inference procedure the model can be extended to work with multiple measurements at different points in time. The source contribution and fractionation weights are assumed to be functions with respect to time $\mathbf{f}(\tau)$ and $r(\tau)$ and the measurements correspond to samples in time $X_t = X(\tau_t)$ at discrete time points $\tau_1, \ldots, \tau_N$.

Now the measurements can be grouped into a measurement matrix $\mathbf{X} := [X_1 \cdots X_N] \in \mathbb{R}^{d \times N}$ where the time dimension is along the matrix rows. Inference of the parameters can be done at the identical time points $\mathbf{f}_t = \mathbf{f}(\tau_t)$ and $r_t = r(\tau_t)$ such that they can be grouped into similar

matrices as well $\mathbf{F} := [\mathbf{f}_1 \cdots \mathbf{f}_N] \in \mathbb{R}^{K\times N}$ and $\mathbf{r} = [r_1 \cdots r_N] \in \mathbb{R}^{1\times N}$. This grouping has the advantage that the mixing equation can be expressed in vectorized form over all time points without changing its general layout.

$$\mathbb{E}[\mathbf{X}|\mathbf{F},\mathbf{r}] = \mu(\mathbf{F},\mathbf{r}) = \mathbf{S}\mathbf{F} + A\log\mathbf{r} \tag{3.4}$$

Again, Rayleigh fractionation is used as representative example, but the mixing equation can be vectorized for other forms of fractionation as well. If multiple consumption pathways exist the corresponding parameters can be expanded to a matrix with L rows $\mathbf{R} \in \mathbb{R}^{L\times N}$ and inference can be done on all $K + L$ parameters per point in time simultaneously.

3.2.1. Independent Time Steps

A simple way to extend the stationary model is to assume complete independence between all points in time. This reduces the time series problem to a set of N independent stationary problems with one single measurement point each and thus the same stationary FRAME model can be used for each point. The vector of measurement errors $\eta \in \mathbb{R}^d$ is now also allowed to vary in time $\eta_1, \ldots, \eta_N$.

$$\begin{aligned} \mathbf{f}_t &\sim \mathrm{Dir}(1), \quad r_t \sim \mathrm{Uni}(0,1) \quad \forall t \\ X_t|\mathbf{f}_t, r_t &\sim \mathcal{N}_d\big(\mu(\mathbf{f}_t, r_t), \eta_t^2\big) \quad \forall t \end{aligned} \tag{3.5}$$

The prior on the series of source contributions $\mathbf{f}_t$ and fractionation weight r_t is now fully independent in time and the information contained in the fact that some measurements are closer in time than others is ignored.

Prior information can be encoded into the prior distribution for $\mathbf{f}_t$ by introducing a concentration parameter $\boldsymbol{\sigma} \in \mathbb{R}_+^K$ as well as a parameter $\alpha \in (0,1)$ that interpolates between the uniform prior for r_t and the Jeffreys prior using the beta distribution as described in subsection 3.1.1. This allows for the inclusion of information that is universally true for all time points simultaneously. If no information is available the model can be extended by adding an additional hierarchical layer for these parameters with weakly informative hyperpriors being the gamma distribution $\Gamma(2,2)$ on the positive real axis for concentrations $\boldsymbol{\sigma}$ and the uniform $\mathrm{Uni}(0,1)$ for α.

$$\begin{aligned} \boldsymbol{\sigma} &\sim \Gamma(2,2), \quad \alpha \sim \mathrm{Uni}(0,1) \\ \mathbf{f}_t &\sim \mathrm{Dir}(\boldsymbol{\sigma}), \quad r_t \sim \mathrm{B}(\alpha,1) \quad \forall t \\ X_t|\mathbf{f}_t, r_t &\sim \mathcal{N}_d\big(\mu(\mathbf{f}_t, r_t), \eta_t^2\big) \quad \forall t \end{aligned} \tag{3.6}$$

3.2.2. Gaussian Process Priors

Time series information can be incorporated by various methods in the case of unconstrained random variables. Since this setting is applicable to the time series of isotopic measurements X_t it is natural to apply the FRAME model with independent time steps on preprocessed measurement series. This preprocessing does not necessarily need to be part of the model itself, but can be done beforehand without consideration for the Bayesian mixing model. Candidate preprocessing algorithms are kernel smoothing, spline smoothing as well as local polynomial regression.

While the latter can offer uncertainty estimates of the smoothed time series, a holistic treatment of estimation with uncertainty can be offered by Gaussian process regression.

Despite the possibility of running the simple time-independent model on preprocessed measurement time series, it can be beneficial to combine both steps into an advanced model. This is due to that fact that for instance the problem specific geometry can influence the feasibility of a region in measurement space. A combined model will include a Gaussian process prior on the measurements X_t such that posterior means W_t can be estimated and used to plug into the above constructed FRAME model with independent time steps. The Gaussian process is shifted and scaled to align with the empirical mean $\hat{\mu}_X$ and standard deviation $\hat{\sigma}_X$ of the measurements X_t and controlled by a kernel function $\mathbf{G}$.

$$\begin{aligned} \mathbf{f}_t \sim \mathrm{Dir}(1), \quad r_t \sim \mathrm{Uni}(0,1) \quad \forall t \\ W_t | \mathbf{f}_t, r_t \sim \mathcal{N}_d\big(\mu(\mathbf{f}_t, r_t), \frac{\eta_t^2}{2}\big) \quad \forall t \\ \mathbf{W} \sim \mathcal{GP}_d(\hat{\mu}_X, \hat{\sigma}_X, \mathbf{G}) \\ X_t | W_t, \mathbf{f}_t, r_t \sim \mathcal{N}_d\big(W_t, \frac{\eta_t^2}{2}\big) \quad \forall t \end{aligned} \tag{3.7}$$

The distribution on the latent estimates W_t is the product of the Gaussian process prior as well as the independent normal distribution around the mixing estimate. Ideally, this model does not need to include sampling of $\mathbf{f}_t$ and r_t because if the mixing equation can be expressed as a linear system of equations (Equation 2.14) then the smooth measurement series W_t is sufficient to solve for the source contribution and fractionation parameters directly. In practice, this approach reduces to applying isotopic mapping techniques to the time series that is preprocessed using Gaussian process smoothing.

If the mixing equation is not explicitly inverted but evaluated by sampling the parameters $\mathbf{f}_t$ and r_t, then the latent variables $\mathbf{W}$ can be marginalized over and eliminated from the model. The product density of $\mathbf{W}$ can also be expressed using known identities [90, ch. 8.1.8] for each separate isotopic measurement dimension $j = 1, \ldots, d$ in terms of its empirical mean $\hat{\mu}_{X,j}$, empirical standard deviation $\hat{\sigma}_{X,j}$ and noise variance η_j^2.

$$\tilde{\Sigma}_j := \left(\frac{2}{\eta_j^2} \mathbb{1} + \frac{1}{\hat{\sigma}_{X,j}^2} \mathbf{G}^{-1} \right)^{-1} \in \mathbb{R}^{N \times N}, \quad \tilde{\mu}_j := \tilde{\Sigma}_j \left(\frac{2}{\eta_j^2} \mu_j^{\mathrm{T}}(\mathbf{F}, \mathbf{r}) + \frac{\hat{\mu}_{X,j}}{\hat{\sigma}_{X,j}^2} \mathbf{G}^{-1} \mathbf{1} \right) \in \mathbb{R}^N$$

Using the Cholesky decomposition these distribution parameters can efficiently be computed and used for sampling. Thus the latent parameters $\mathbf{W}$ can be eliminated from the model and the likelihood of each row $\mathbf{X}_{j:}$ can be directly computed.

$$\begin{aligned} \mathbf{F} \sim \mathrm{Dir}(1), \quad \mathbf{r} \sim \mathrm{Uni}(0,1) \\ \mu(\mathbf{F}, \mathbf{r}) \sim \mathcal{GP}_d(\hat{\mu}_X, \hat{\sigma}_X, \mathbf{G}) \\ \mathbf{X}_{j:}^{\mathrm{T}} | \mathbf{F}, \mathbf{r} \sim \mathcal{N}_N\big(\tilde{\mu}_j, \tilde{\Sigma}_j + \frac{\eta_t^2}{2} \mathbb{1}\big) \quad \forall j = 1, \ldots, d \end{aligned} \tag{3.8}$$

Gaussian process priors on measurements use only one single hyperparameter which is the correlation length ρ used to compute the kernel matrix $\mathbf{G}_{ij} = \kappa_\rho(\tau_i, \tau_j)$. The scale of the Gaussian process is always set to the empirical standard deviation of the data and is thus considered fixed. In order to compile a fully hierarchical Bayesian model an inverse gamma distribution $\frac{1}{\rho} \sim \Gamma(2, 2)$ can be used as hyperprior for the correlation length assuming that the time scales are properly normalized.

3.2.3. Generalized Gaussian Process Priors

To make use of time series information in the source contribution and fractionation weights direct priors are desired. Such priors can be constructed by sampling auxiliary variables from multiple independent Gaussian processes $\mathbf{Z} \sim \mathcal{GP}_K(\mathbf{G})$ and at each point in time inverting the log ratio transformations on the simplex introduced in section 2.5 in order to create a time series of simplex-valued variables $\mathbf{f}_t$. The fractionation amount r_t is constrained to the interval $[0, 1]$ and can thus be linked for instance by applying the logit transform $\text{logit}(r) = \log \frac{r}{1-r}$ at each point in time, which maps it to the entire real axis. Hyperparameters for correlation length $\rho \in \mathbb{R}_+$ and concentration $\sigma \in \mathbb{R}_+$ are used to compute the kernel matrix $\mathbf{G}_{ij} = \sigma^2 \kappa_\rho(\tau_i, \tau_j)$ for the Gaussian process. The general shape of these priors is visualized in Figure 3.1 and Figure 3.2. Working with the matrix of source contributions $\mathbf{F} = [\mathbf{f}_1 \dots \mathbf{f}_N] \in \mathbb{R}^{K \times N}$ and fractionation weight $\mathbf{r} = [r_1 \cdots r_N] \in \mathbb{R}^{1 \times N}$ the model can be stated in vectorized form, where the link functions are understood to be column-wise.

$$\begin{aligned} \text{clr}(\mathbf{F}) &\sim \mathcal{GP}_K(\mathbf{G}) \\ \text{logit}(\mathbf{r}) &\sim \mathcal{GP}(\mathbf{G}) \\ X_t | \mathbf{f}_t, r_t &\sim \mathcal{N}_d\left(\mu(\mathbf{f}_t, r_t), \eta_t^2\right) \quad \forall t \end{aligned} \tag{3.9}$$

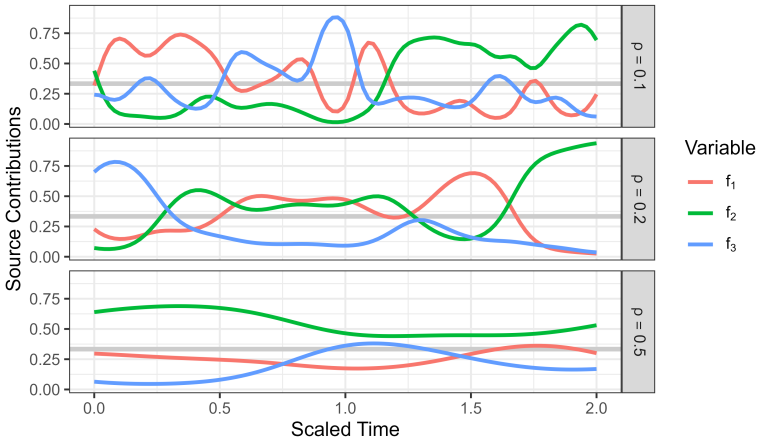

Figure 3.1.: Gaussian process prior with radial basis function kernel and with different correlation lengths ρ for three source contributions mapped to the simplex using the centered log ratio transform.

Both link functions used can easily be inverted once random variables $\mathbf{Z} \sim \mathcal{GP}_K(\mathbf{G})$ and $\mathbf{Y} \sim \mathcal{GP}(\mathbf{G})$ are sampled from Gaussian processes over time $t = 1, \ldots, N$. The inverse of the clr transform is given by the softmax function and the inverse of the logit link is given by the sigmoid function.

$$\mathbf{F}_{kt} = \mathrm{clr}_k^{-1}(Z_{1t}, \ldots, Z_{Kt}) = \frac{e^{Z_{kt}}}{\sum_{j=1}^{K} e^{Z_{jt}}}, \qquad r_t = \mathrm{logit}^{-1}(Y_t) = \frac{1}{1 + e^{-Y_t}}$$

The prior on the source contribution parameters $\mathbf{F}$ is known as a generalized Gaussian process prior and techniques such as Taylor expansion can be used to derive analytic approximations [91]. Its marginal is a softmax transformed multivariate Gaussian, which is also known as a logistic-normal distribution and serves as an approximation to the Dirichlet distribution [92]. This is due to the fact that a Dirichlet distribution can be constructed as the ratio of gamma distributed variables [93, ch. 4.1] and the softmax function produces a ratio of log-normal variables $e^{Z_{kt}}$ that are very similar in distribution. The mapping using centered log ratio transforms thus creates a time series of random variables with approximate Dirichlet marginals, which is referred to as a *Dirichlet-Gaussian process (DGP)* [94].

The marginals are controlled by the parameter σ of the Gaussian process that now acts as the concentration parameter of the Dirichlet distribution. Since the covariance kernel $\mathbf{G}$ is scaled to generate Gaussian random variables with unit variance if $\sigma = 1$, the marginal distribution in that case is approximately the uniform $\mathrm{Dir}(1)$. This fact can be seen by sampling from these generalized Gaussian process priors and estimating the marginals as in Figure 3.2.

Using the isometric log ratio transform ilr instead of clr reduces the number of Gaussian processes that need to be sampled to $K - 1$ for the source contributions. Inverting this link function is as simple as applying an orthonormal base transform $\mathbf{U}$ (section 2.5) to the random variables and then applying the softmax function. Since interpretability of the sampled Gaussian process variables is not required, any orthonormal basis is suitable and a simple construction using Gram-Schmidt orthogonalization is chosen [84].

$$\begin{aligned} \mathrm{ilr}(\mathbf{F}) &\sim \mathcal{GP}_{K-1}(\mathbf{G}) \\ \mathrm{logit}(\mathbf{r}) &\sim \mathcal{GP}(\mathbf{G}) \\ X_t | \mathbf{f}_t, r_t &\sim \mathcal{N}_d\big(\mu(\mathbf{f}_t, r_t), \eta_t^2\big) \quad \forall t \end{aligned} \tag{3.10}$$

Shorthand notation $\mathcal{GP}(\mathbf{G}) = \mathcal{GP}(0, 1, \mathbf{G})$ is used with correlation length ρ and scale σ included in the kernel computation $\mathbf{G}_{ij} = \sigma^2 \kappa_\rho(\tau_i, \tau_j)$. Both kernel parameters ρ and σ can be set in advance or given weak hyperpriors. The inverse gamma distribution $\frac{1}{\rho} \sim \Gamma(2, 2)$ and the regular gamma distribution $\sigma \sim \Gamma(2, 2)$ are chosen under the assumption that the time variables $\tau_1, \ldots, \tau_N$ are scaled appropriately. This hierarchical model benefits especially from the reduced number of Gaussian processes sampled when using the ilr transform since the kernel covariance matrix must be reconstructed in every sampling step. The number of hyperparameters can be increased by using separate concentrations and correlation lengths for the source contributions $\mathbf{F}$ and the fractionation weights $\mathbf{r}$.

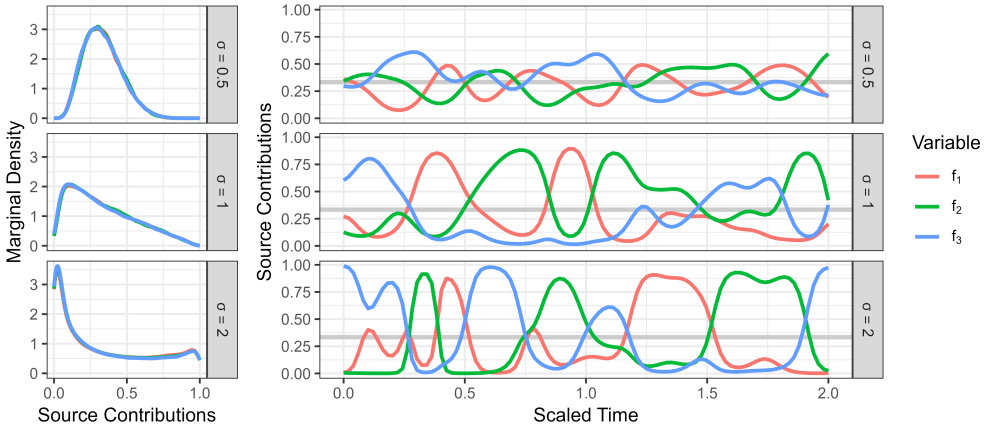

Figure 3.2.: Estimated marginal densities of transformed Gaussian process priors for different concentration parameters σ on the left together with one prior observation for three source contributions on the right.

3.2.4. Spline-Based Priors

An alternative to Gaussian process priors are spline basis functions, which can be used to construct a linear fitting operation that is then mapped to simplex space. This allows for the addition of exogenous variables as predictors of source contributions or fractionation as well. A cubic spline basis of M basis functions is evaluated at the measurement points $\tau_1, \ldots, \tau_N$ to form the evaluation matrix $\mathbf{H} \in \mathbb{R}^{N \times M}$ with $\mathbf{H}_{ij} = s_j(\tau_i)$ for polynomial basis functions $s_1(\cdot), \ldots, s_M(\cdot)$. The time series of source contributions in simplex space is reconstructed with the basis coefficients $\mathbf{b}_k \in \mathbb{R}^M$ for each source $k = 1, \ldots, K$ arranged to the matrix $[\mathbf{b}_1 \cdots \mathbf{b}_K]^\mathrm{T} = \mathbf{B} \in \mathbb{R}^{K \times M}$ and coefficients for fractionation $\mathbf{c} \in \mathbb{R}^{1 \times M}$. This type of model is therefore part of the generalized linear model class [95] and allows for easy extension with fixed effects relating to measurement dimensions as well as random effects for experiment replication. It will thus further be referred to as *generalized linear model with spline basis (spline GLM)*.

$$\begin{aligned} \mathbf{B}, \mathbf{c} &\sim \mathcal{N}(0, 1) \\ \operatorname{clr}(\mathbf{F}) &= \mathbf{B}\mathbf{H}^\mathrm{T} \\ \operatorname{logit}(\mathbf{r}) &= \mathbf{c}\mathbf{H}^\mathrm{T} \\ X_t | \mathbf{f}_t, r_t &\sim \mathcal{N}_d\left(\mu(\mathbf{f}_t, r_t), \eta_t^2\right) \quad \forall t \end{aligned} \tag{3.11}$$

In consequence, the distribution of the basis coefficient vector $\mathbf{b}_k = \mathbf{B}_{k:}^\mathrm{T} \in \mathbb{R}^M$ before transformation has distribution $\mathbf{b}_k \sim \mathcal{N}_M(0, \mathbb{1})$ for source $k = 1, \ldots, K$. After application of the spline basis transform it is thus still Gaussian $\mathbf{H}\mathbf{b}_k \sim \mathcal{N}_N(0, \mathbf{H}\mathbf{H}^\mathrm{T})$ although with a modified covariance matrix $\mathbf{H}\mathbf{H}^\mathrm{T} \in \mathbb{R}^{N \times N}$. Since the inverse centered log ratio transform maps Gaussian random variables with unit variance approximately to a uniform Dirichlet distribution, it makes sense to scale the basis transform such that $\frac{1}{N}\mathrm{Tr}(\mathbf{H}\mathbf{H}^\mathrm{T}) = 1$ as the spline basis vectors are not semi-orthogonal in general.

The linear estimator described above with Gaussian priors on its coefficients is firmly linked to a ridge regression estimator in frequentist statistics, since both the posterior mean and the maximum a posteriori solution coincide with the ridge estimator [96, ch. 6.2.1]. This estimator is used to introduce shrinkage into the parameter values, thereby regularizing the effect of an excessively large number of degrees of freedom. Furthermore, it can vaguely be connected to smoothing splines by restricting the basis to natural cubic splines and using a generalized version of ridge regression with weighted data points [75], [97], [98].

Similarly to Dirichlet-Gaussian process prior models, using the isometric log ratio transform reduces the number of spline basis coefficient vectors needed for the source contributions to $K-1$. However, in this case the benefit is smaller since the coefficient vectors have only size $M < N$. Further alterations can be made by using a different number of degrees of freedom between source contributions and fractionation. The basis coefficient priors are also a way of customizing the model by scaling them to standard deviation $\sigma \in \mathbb{R}_+$ and thus introducing a concentration parameter or by setting the prior shape to a Laplacian distribution, which has heavier tails and is linked to the frequentist LASSO regression estimator, albeit only for the maximum a posteriori solution [49], [96].

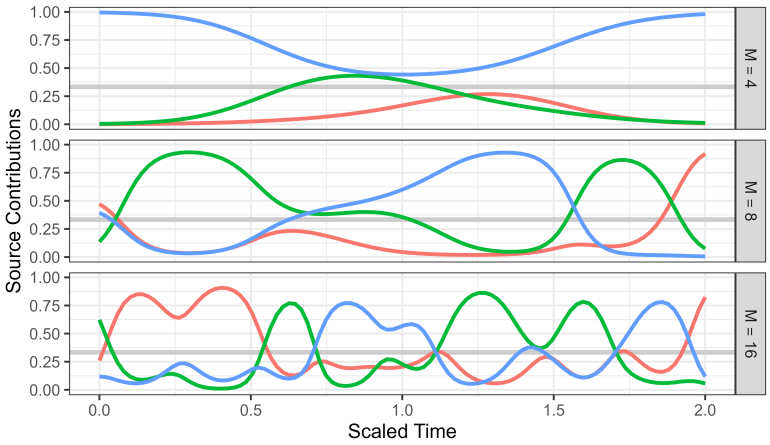

Figure 3.3.: Spline prior for three source contributions transformed to the simplex with the clr transformation using different degrees of freedom M that can be used to control the covariance of source contributions at separate points in time.

3.3. Model Comparison

3.3.1. Data Simulation

Models can be compared by simulating the data generating process multiple times and then comparing the resulting posterior sample with fixed truth values. The time series of source contribution and fractionation weights used to simulate the data are denoted by $\mathbf{F}^* = [\mathbf{f}_1^* \cdots \mathbf{f}_N^*]$ and $\mathbf{r}^* = [r_1^* \cdots r_N^*]$ and the mixing equation using Rayleigh fractionation is chosen. Measurement

generation is then repeated Q times by sampling the source isotopic signature $\mathbf{S}^{(q)} \in \mathbb{R}^{d\times K}$ and fractionation factors $A^{(q)} \in \mathbb{R}^d$ from their respective priors and then adding independent Gaussian measurement errors $E_t^{(q)} \sim \mathcal{N}_d(0, \eta^2)$ with noise variance $\eta^2 \in \mathbb{R}^d$ for $q = 1, \ldots, Q$.

$$X_t^{(q)} = \mathbf{S}^{(q)}\mathbf{f}_t^* + A^{(q)} \log r_t^* + E_t^{(q)} \tag{3.12}$$

Auxiliary data for source isotopic signature and fractionation factor used to generate them are taken from Yu, Harris, Lewicka-Szczebak, *et al.* [99]. They correspond to nitrification (S_1) and bacterial denitrification (S_2). Priors are uniform for the sources $S_j \sim \text{Uni}(b_j, \Delta_j)$, $j = 1, 2$ and Gaussian for the fractionation factor with variance matched to the reported bounds $A \sim \mathcal{N}(c, \frac{\Delta_A^2}{12})$.

	$b_1 \pm \Delta_1$	$b_2 \pm \Delta_2$	$c \pm \Delta_A$
$\delta^{15}\text{N}$	-55.5 ± 17.0	-25.25 ± 55.1	-6.4 ± 9.2
$\delta^{15}\text{N}^{\text{SP}}$	35.35 ± 6.7	-1.9 ± 11.2	-5.55 ± 5.3

Table 3.1.: Prior distribution parameters for two source isotopic signatures and one fractionation factor taken from Yu, Harris, Lewicka-Szczebak, *et al.* [99] that are used to simulate data sets for model testing.

The comparisons are done on fixed parameter values that intend to be illustrative for a given edge case that might occur in reality. The true parameter time series $\mathbf{f}_t^*$ and r_t^* are shown in Figure 3.4 and are sampled at $N = 32$ equally spaced time steps. One additional general example (GenE) is used for simulation with properties being less extreme than for the other ones and hopefully representative of average datasets that might be encountered in practice. It is used for subsection 4.3.3 and subsection 4.3.4 and studied extensively in section 4.4.

Bayesian parameter estimation is then tested on each generated data set $\mathbf{X}^{(q)} = [X_1^{(q)} \cdots X_N^{(q)}] \in \mathbb{R}^{d\times N}$ for $q = 1, \ldots, Q$ individually and a total of S posterior samples of all parameters is produced each time. The posterior samples shall be denoted by $\mathbf{F}^{(q,s)} = [\mathbf{f}_1^{(q,s)} \cdots \mathbf{f}_N^{(q,s)}]$ and $\mathbf{r}^{(q,s)} = [r_1^{(q,s)} \cdots r_N^{(q,s)}]$ respectively for $s = 1, \ldots, S$.

All experiments are run on an Intel Core i9-10900K CPU. The reported runtimes in section 4.2 are of a single sampling chain and in section 4.3 the reported times are the maximum of 4 simultaneously run chains.

3.3.2. Measuring Quality of Inference

Sampling from the posterior distribution does not give unique point estimates for the parameters involved and multiple ways of computing them exist. Most commonly the posterior mean is used as point estimate, although using the median for example could also be a useful strategy for posterior distributions that are highly dissimilar to a Gaussian distribution.

$$\hat{\mathbf{F}}^{(q)} = \hat{\mathbb{E}}\left[\mathbf{F}\middle|\mathbf{X}^{(q)}\right] = \frac{1}{S}\sum_{s=1}^{S}\mathbf{F}^{(q,s)}, \quad \hat{\mathbf{r}}^{(q)} = \hat{\mathbb{E}}\left[\mathbf{r}\middle|\mathbf{X}^{(q)}\right] = \frac{1}{S}\sum_{s=1}^{S}\mathbf{r}^{(q,s)} \tag{3.13}$$

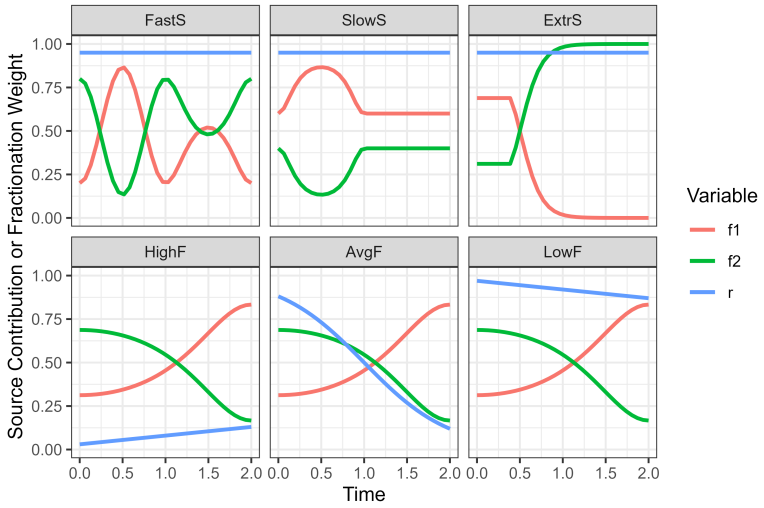

Figure 3.4.: True parameter series used to simulate datasets that illustrate fast changing source contributions (FastS), slow changing source contributions (SlowS), extremal source contributions (ExtrS), high fractionation (HighF), average and fast changing fractionation (AvgF) and low fractionation (LowF).

Now the accuracy of the estimation can be assessed by computing the distance between these pointwise estimates and the true value such as root mean squared error (RMSE) and mean average error (MAE). Although it is possible to compute the metrics at each point in time they are averaged for simpler model comparison. Computations for $\mathbf{r}^{(q,s)}$ are analogous.

$$\mathrm{RMSE}_k^{(q)} := \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\hat{\mathbf{F}}_{kt}^{(q)} - \mathbf{F}_{kt}^{*}\right)^2}, \quad \mathrm{MAE}_k^{(q)} := \frac{1}{N}\sum_{t=1}^{N}\left|\hat{\mathbf{F}}_{kt}^{(q)} - \mathbf{F}_{kt}^{*}\right| \tag{3.14}$$

Since the parameters to be estimated are interpreted as a time series, it makes sense to also compare specific time series information across the model estimates. The rate of change can significantly be confounded in the measurement time series, since the measurement errors follow a white noise distribution that introduces high frequency changes. The ability of models to filter this noise can be measured by comparing the rate of change which is approximated using first differences $\Delta\mathbf{F}_{tj}^{*} = \mathbf{F}_{t+1,j}^{*} - \mathbf{F}_{tj}^{*}$ and $\Delta r_t^{*} = r_{t+1}^{*} - r_t^{*}$ for $t = 1, \ldots, N-1$. The magnitude of changes is not necessarily relevant, since misestimation of them would also lead to bad pointwise estimates. Therefore the ratio of variances of first differences shall serve as comparison metric, which can be understood as comparing a notion of curvature or acceleration.

$$\mathrm{VarFD}_k^{(q)} = \frac{\sum_{t=1}^{N-1}\left(\Delta\mathbf{F}_{kt}^{(q)} - \frac{1}{N}\sum_{j=1}^{N-1}\Delta\mathbf{F}_{kj}^{(q)}\right)^2}{\sum_{t=1}^{N-1}\left(\Delta\mathbf{F}_{kt}^{*} - \frac{1}{N}\sum_{j=1}^{N-1}\Delta\mathbf{F}_{kj}^{*}\right)^2} \tag{3.15}$$

3.3.3. Metrics for Bayesian Posteriors

Bayesian models are mainly used to derive pointwise estimates, but their advantage is the creation of a sample from the posterior distribution. It is thus also important to take distribution properties into account. A simple way of doing so is evaluating the estimated posterior density at the true parameter value, that is known from the data simulation. High posterior density values provide evidence that maximum a posteriori estimates could be appropriate.

Equally-tailed credible intervals can be constructed using the estimated posterior quantiles $q(\gamma|\mathbf{X})$, although different notions of the credible interval such as highest density intervals exist [100]. The credible interval with level $\gamma \in [0, 1]$ is then the set excluding the tails with a proportion of $\frac{\gamma}{2}$ of the most extreme observations on either side.

$$I(\gamma) = \left[q\left(\frac{1-\gamma}{2}\middle|\mathbf{X}\right), q\left(\frac{1+\gamma}{2}\middle|\mathbf{X}\right)\right]$$

Posterior interval coverage is a useful metric for simulated data of the full Bayesian model, meaning in particular that $\mathbf{F}^*$ and $\mathbf{r}^*$ are sampled from their priors as well. Since the data simulation for model comparison used here has fixed parameter values, interpreting interval coverage becomes less meaningful. It is still practical to use the size of the credible interval as measure of uncertainty and thus the interval span $q\left(\frac{1+\gamma}{2}\middle|\mathbf{X}\right) - q\left(\frac{1-\gamma}{2}\middle|\mathbf{X}\right)$ can be compared, which is desired to be as small as possible given otherwise accurate estimates.

Further metrics for the quality of the entire posterior distributions can be taken into consideration. Posterior predictive checks are typically used in cases where no true values for the parameter estimates are available in order to assess the models capability of representing the input data well [101]. The posterior predictive distribution is the likelihood of hypothetical future measurements calculated using the posterior distribution over parameter values in presence of the actually available data.

$$p(\tilde{\mathbf{X}}|\mathbf{X}) = \int p(\tilde{\mathbf{X}}|\mathbf{F}, \mathbf{r})\pi(\mathbf{F}, \mathbf{r}|\mathbf{X})d\mathbf{F}d\mathbf{r} \tag{3.16}$$

This posterior predictive distribution is not unique since given that there are auxiliary parameters $\mathbf{S}$ and A as well it is unclear whether it should be proportional to the marginalized likelihood $p(\tilde{\mathbf{X}}|\mathbf{F}, \mathbf{r})$ or rather the likelihood conditioned on the auxiliary parameters $p(\tilde{\mathbf{X}}|\mathbf{F}, \mathbf{r}, \mathbf{S}, A)$. This discrepancy renders comparison of predictive density values across dataset simulations $\mathbf{X}^{(q)}$ ineffective, since the models fit a joint posterior and thus assume that future data must be sampled using identical auxiliary parameter values, whereas the simulation resamples their values $\mathbf{S}^{(q)}, A^{(q)}$ every time.

The log pointwise predictive density is a metric for the quality of the posterior predictive density and thus by proxy of the Bayesian model. It is computed by evaluating the predictive density at the original data points and can be approximated using the posterior samples [102].

$$\text{LPPD}^{(q)} = \log \int p(\mathbf{X}|\mathbf{F}, \mathbf{r})\pi(\mathbf{F}, \mathbf{r}|\mathbf{X})d\mathbf{F}d\mathbf{r} \approx \frac{1}{S}\sum_{s=1}^{S} p(\mathbf{X}|\mathbf{F}^{(s)}) \tag{3.17}$$

Using the model fitted on the same data as it is evaluated tends to overestimate the pointwise predictive density and new data is assumed to have lower values in reality. Ideally, the expectation over the true distribution of the data is desired for comparison, which is not feasible in general since the marginal $p(\mathbf{X})$ is not known. Approximations to out-of-sample predictive densities exist with prominent approaches being the Watanabe-Akaike information criterion or leave-one-out cross-validation [103]. However, they assume that multiple independent draws are available and used to fit the model, which is not the case for the data used to compare models. Even though datasets encountered in practice reflect this choice, it is not inconceivable that real datasets can be obtained where multiple time series are at hand, that can be treated as independent observations and thus allow for the computation of these approximations to the expected log predictive density.

Chapter 4

Results & Discussion

4.1. Rayleigh Fractionation Priors

Sampling the posterior distribution of the fractionation weight r for closed system fractionation is hard, because it is connected through the non-linear logarithm to the effect on measurements. Although a uniform prior usually does not inform the posterior about anything else than the boundaries, the logarithm makes its effect on the posterior much more unclear. Different choices for prior distributions are thus tested for their effect on the generated posterior sample.

The simulated dataset uses 17 different values for the fractionation index r ranging from 0.05 up to 0.95. Source contributions are fixed to $f_1^* = 0.7$ and $f_2^* = 0.3$. Each value of r is used to generate $Q = 64$ data points $X^{(1)}, \ldots, X^{(Q)}$ with measurement error $\eta = 4$ for a total of 1088 data points. The stationary inference model given in Equation 3.1 is fitted to each point individually, which makes this setting analogous to the inference procedure used in the original FRAME model [18].

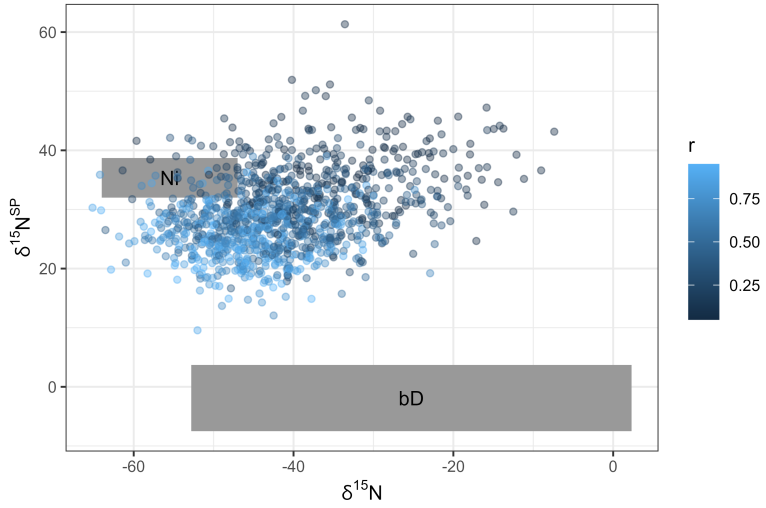

Figure 4.1.: Dual isotope plot for the two sources used with simulated data points colored by the fractionation index r.

Using a uniform prior for the fractionation index $\pi(r) \propto 1$ is the natural choice used as standard by all models. It is compared to the Jeffreys prior, which was derived in subsection 3.1.1 to be $\pi(r) \propto \frac{1}{r}$ and thus is an improper prior. A middle ground between these two choices is given

by the beta prior $r \sim B(\frac{1}{2}, 1)$, which has $\pi(r) \sim \frac{1}{\sqrt{r}}$. The first argument of the beta distribution can also be used as a free parameter $\alpha \in (0, 1]$ alongside a uniform hyperprior to construct the hierarchical model $r \sim B(\alpha, 1)$, $\alpha \sim \text{Uni}(0, 1)$.

Each data point is supplied to the model individually and $S = 5000$ posterior samples are generated using Stan. The samples are combined for each value of r to marginalize over the distributions of the auxiliary parameters. Posterior density estimations for $r = 0.05$, $r = 0.5$ and $r = 0.95$ are shown in Figure 4.2.

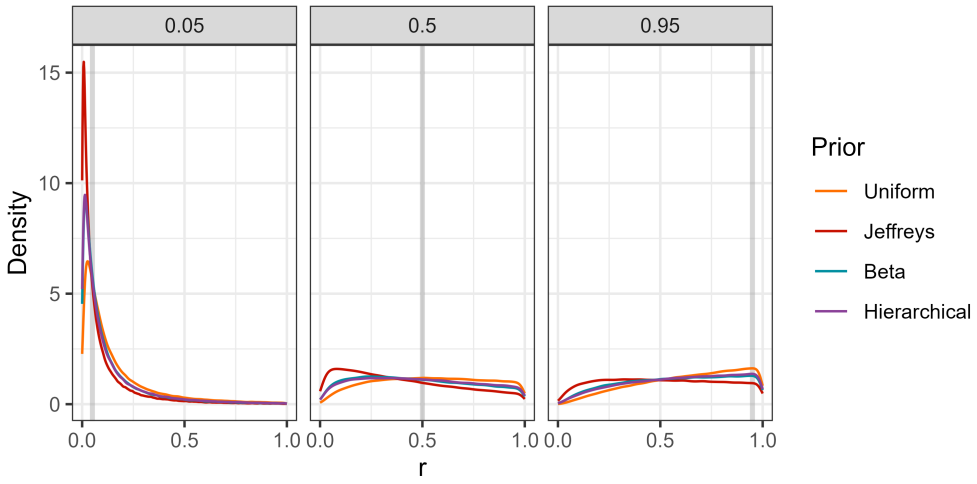

Figure 4.2.: Posterior densities of the fractionation weight r averaged across simulations for different prior distributions.

The source S_2 (bD) has a wide prior distribution which confounds with the effect of fractionation and introduces high uncertainty. Since this uncertainty is proportional to the logarithm of r, the effect on the posterior distribution is much more pronounced when r is high. The Jeffreys prior introduces a shift towards lower values compared to the uniform prior with the beta prior being in between the two. The hierarchical prior estimates a parameter value of $\alpha \approx \frac{1}{2}$ closely. This can be understood as a maximum evidence estimator since the prior on α was uniform, and consequently the beta prior represents a distribution derived from empirical Bayesian analysis.

Inference performance is compared by taking the posterior means $\hat{\mathbf{f}}^{(q)} = \frac{1}{S}\sum_{s=1}^{S} \mathbf{f}^{(q,s)}$ and $\hat{r}^{(q)} = \frac{1}{S}\sum_{s=1}^{S} r^{(q,s)}$ and comparing them against the true values $\hat{\mathbf{f}}^{*}$ and r^{*}. Figure 4.3 shows the mean absolute error of the posterior mean with vertical lines indicating the standard deviation over the $Q = 64$ repetitions. Clearly, the Jeffreys prior performs worst for source contributions with the uniform prior being best and closely followed from the beta and hierarchical prior. In terms of fractionation index the Jeffreys priors is best for low values of r with the uniform being worst, but this relationship switches at about $r = 0.4$ to the contrary. Therefore, choosing any prior can still be justified if one expects the fractionation index to be in a certain range and the effect of prior choice is overwhelmed by the variation introduced through the distribution of the sources regardless.

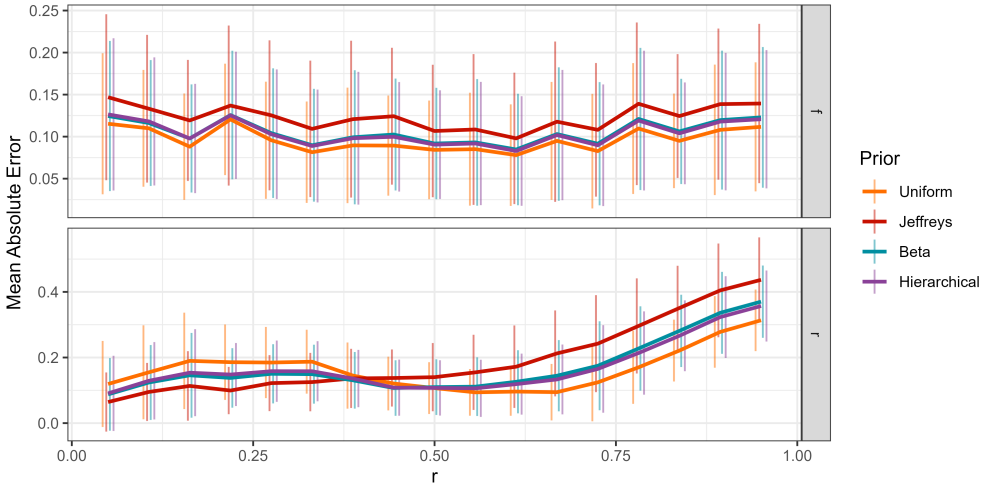

Figure 4.3.: Mean absolute error of Bayesian models using different fractionation prior distributions over different true fractionation values. The performance on source contributions is identical for both sources, since they are perfectly correlated, so only one panel is shown for both.

4.2. Sampling Software

Bayesian models can be implemented using different sampling strategies. The most commonly used sampling libraries are JAGS using Gibbs sampling and Stan using Hamiltonian Monte Carlo sampling. These two implementations are compared in three different settings. Firstly, the original stationary FRAME estimator for one single time step given in Equation 3.1, then the time series model using independent time steps described in Equation 3.5 and the hierarchical Dirichlet-Gaussian process model described in Equation 3.9, with joint estimation of the correlation length for source contributions and fractionation.

The general example GenE discussed in section 4.4 serves as fixed truth value for $\mathbf{f}_t^*$ and r_t^* for simulating data and applying the models in the above mentioned three settings. The time series are sampled with $N = 64$ points and for the stationary case the average over time is chosen as fixed value. Each implementation is run for $S = 10000$ sampling steps and the resulting effective sample size, total runtime in seconds and resulting effective samples per second are noted in Table 4.1. The effective sample sizes are computed using the calculation described by Kruschke [104, ch. 7.5.2.] and the reported number are averages over all parameters.

JAGS seems to outperform Stan for the stationary and independent case by having more effective samples in the shorter amount of time. The hierarchical model could only be efficiently sampled by Stan although it took a comparatively long time.

These findings are consistent with results on linear models using different numbers of parameters reported by Beraha, Falco, and Guglielmi [105]. It seems like Stan is a good option for time series models that inherently have a lot of parameters and the cases where JAGS is faster have sufficiently short sampling times for both libraries.

		N_{eff}	time	N_{eff}/s
stationary	Stan	4482	10	454
	JAGS	5131	0.5	9162
independent	Stan	6350	66	97
	JAGS	6559	22	292
hierarchical	Stan	11538	1477	8
	JAGS	10	1527	-

Table 4.1.: Effective sample size, runtime in seconds and effective samples per second for the Stan and JAGS sampler over $S = 10000$ sampling steps.

4.3. Time Series Models

Time series models are compared for a wide range of scenarios representative of edge cases that might occur in reality. For the experiment, fixed parameter values for source contributions $\mathbf{F}^*$ and fractionation $\mathbf{r}^*$ and datasets are sampled according to the procedure described in section 3.3. The models to be compared are the following:

- Independent time step model described in Equation 3.5
- Hierarchical independent time step model with gamma hyperprior for concentration σ described in Equation 3.6
- Gaussian process prior on measurements with $\rho = 1$ described in Equation 3.8
- Gaussian process prior on measurements with $\rho = 1$ using latent variable formulation described in Equation 3.7
- Gaussian process prior on measurements with inverse gamma hyperprior on ρ
- Gaussian process prior on measurements with with inverse gamma hyperprior on ρ using latent variable formulation
- DGP prior using clr transform and $\rho = 1$, $\sigma = 1$ described in Equation 3.9
- DGP prior using ilr transform and $\rho = 1$, $\sigma = 1$ described in Equation 3.10
- DGP prior using clr transform and $\sigma = 1$ with inverse gamma hyperprior on ρ
- DGP prior using ilr transform and $\sigma = 1$ with inverse gamma hyperprior on ρ
- B-spline GLM using clr link function having $M = 8$ for source contributions and $M = 4$ for fractionation described in Equation 3.11
- B-spline GLM using ilr link function and $M = 8$ for source contributions and $M = 4$ for fractionation
- B-spline GLM using ilr link function and $M = 8$ for source contributions and $M = 4$ for fractionation with Laplace prior on coefficients

4.3.1. Illustrative Examples

The examples described in subsection 3.3.1 are sampled for $Q = 64$ repetitions. Measurements are simulated with Gaussian measurement error of magnitude $\eta = 5$. The posteriors are sampled for a total of $S = 10000$ steps using 4 parallel chains. Goodness of estimation is quantified with estimates computed as the posterior means $\hat{\mathbf{f}}_t^{(q)} = \frac{1}{S}\sum_{s=1}^{S}\mathbf{f}_t^{(q,s)}$ and $\hat{r}_t^{(q)} = \frac{1}{S}\sum_{s=1}^{S} r_t^{(q,s)}$ and taking the mean absolute error to the ground truth. Both the root mean squared error and mean absolute error have been found to be similar and posterior medians lead to the same conclusions as well. Mean absolute error results are shown in Figure 4.4 and details for all metrics are available in Appendix B. The results for the two source contributions were identical since they are perfectly correlated, so the plots only show one result for $\mathbf{f}$ in addition to the fractionation weight r if applicable.

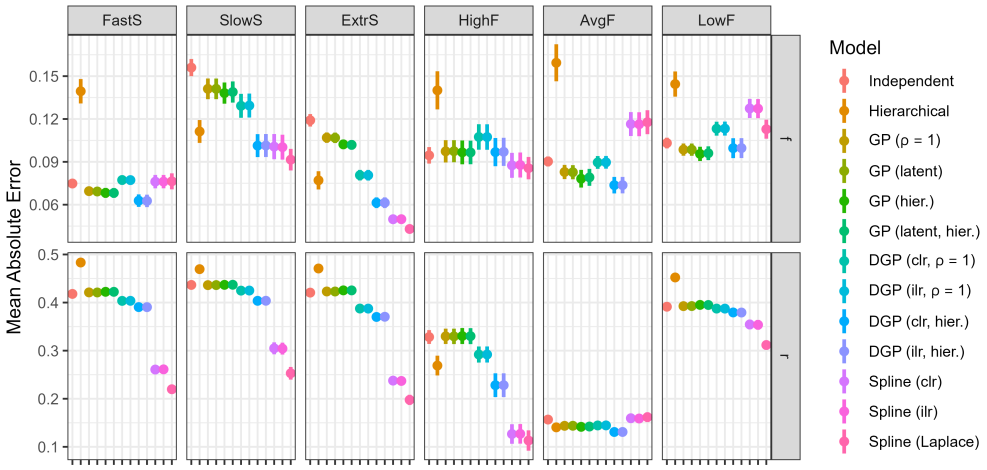

Figure 4.4.: Mean absolute errors for all examples averaged over $Q = 64$ dataset simulations with standard deviations shown as vertical lines.

All models using fixed hyperparameters use default values that are not specifically tuned for the examples at hand. Therefore the reported performance is not indicative of best case performance and only shows the quality of the chosen values. Hierarchical models do not have this problem since they can estimate the hyperparameters for each example specifically.

Overall performance seems to be best for hierarchical DGP models or spline GLMs. Estimation of fractionation is not ideal in most cases with spline models performing best for extremal fractionation amounts, which are all examples except AvgF. The default number of degrees of freedom that spline models use seems surprisingly robust in all examples whereas the default correlation length of Gaussian processes does is not. Gaussian process priors on measurements appear to be slightly worse than DGP priors and spline-based priors especially for examples SlowS and ExtrS. Independent time step models have worse performance than the rest for all examples and the hierarchical extension to it only has good performance in example SlowS and ExtrS, which represent cases where concentration to sources is either very low or very high. In other cases using flat priors as default values seems to work best.

Spline GLMs have very small errors on fractionation whenever the value is slowly changing and close to 0 or 1. This could suggest that the chosen hyperparameters are suitable for all examples. Another plausible explanation is the fact that the spline bases used have an intercept term, which allows it to freely move the center of estimation, whereas DGP models do not. For the source contributions this likely does not matter, but allowing the Gaussian process to have non-zero mean could be beneficial for fractionation. An additional spline model was added with Laplace priors on the coefficients. This seems to be beneficial in cases where parameters are close to their boundaries since the prior allows for values farther from zero.

The model using Gaussian process priors on measurements was implemented using both formulations with latent variables in Equation 3.7 and with analytically computed likelihood in Equation 3.8. Performance is identical in all cases hence strongly indicating that the formulations are equivalent for parameter estimation. Additionally, using clr or ilr transformations for DGP models and Spline GLMs does not make a difference in estimation accuracy, as is to be expected from their derivations.

Time series models are not only expected to give accurate estimates of the source contributions and fractionation, but the resulting time series should have similar properties to the truth as well. The variance ratio of first differences described in Equation 3.15 measures accuracy of the estimated curvature. Results for source contributions are shown in Figure 4.5.

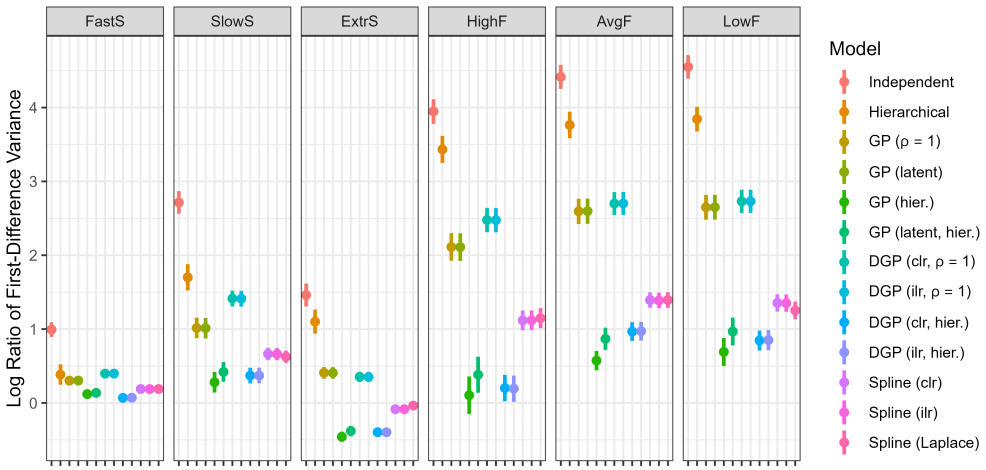

Figure 4.5.: Log variance ratio of first differences of the estimated time series against the true values for source contributions. The time series for fractionation are linear or constant in most examples and are thus not suitable to be used for model comparison.

Clearly, the hierarchical Gaussian process and DGP models estimate the correlation length well, resulting in a time series with similar rates of change than the true values. Spline GLMs perform well, especially for the examples FastS, SlowS and ExtrS. Independent time step models result in high variation which is due to the fact the the measurement noise is not adequately filtered and the fixed correlation length Gaussian processes seem to have misspecified hyperparameters, since they also overestimate the rates of change in the time series.

Fitting times ranged from 20 seconds to 72 seconds on average for the $S = 10000$ posterior samples generated by each model, split into 2500 over 4 chains. The longest time out of these 4 concurrent chains is shown in Figure 4.6. Hierarchical models tend to be slowest due to the additional parameters and repeated matrix decompositions that need to be computed, whereas fixed parameter models, especially independent time steps and spline GLMs, sample fastest. Spline models using the Laplace prior have long fitting times which could indicate that the high parameter values, that are allowed due to weaker regularization of of parameter ranges far from zero, are not sufficiently identifiable resulting in slow sample generation.

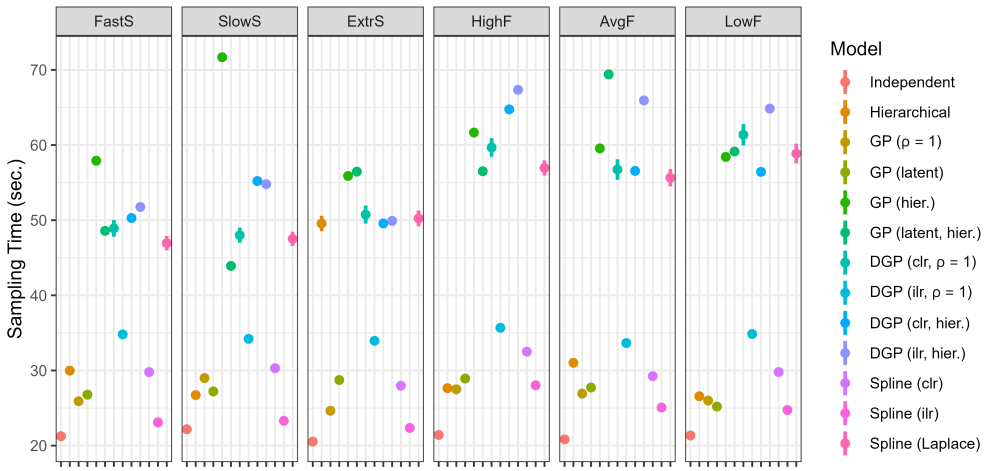

Figure 4.6.: Fitting time in seconds for $S = 10000$ posterior samples split among 4 chains. The reported result is the maximum fitting time among the chains.

4.3.2. Improved Data Quality

Estimation accuracy can be improved not only by choosing the right model but also by improving the data quality. Several ways of adding more or higher quality data exist and the effect on model performance is studied in order to find what would be most beneficial.

The examples above use two sources with two isotopic measurements making the system well-determined. If additional isotopic measurements are available they can be added to make the system overdetermined and thus eliminate some noise. For this the additional isotopic measurement $\delta^{18}O$ is added with source locations and uncertainty reported by Yu, Harris, Lewicka-Szczebak, *et al.* [99]. The same dataset generation procedure as in subsection 4.3.1 is used with a total of $Q = 64$ datasets generated. Resulting improvement in estimation accuracy for the same two sources is shown in Figure 4.7.

An improvement can clearly be seen especially for estimation of the fractionation weight r in the examples HighF and AvgF. It is worth to note that the additional measurement is not ideal in quality having large uncertainty for the fractionation factor. Spline GLMs seem to have improved the most, especially in their already good ability to estimate fractionation weights.

	$b_1 \pm \Delta_1$	$b_2 \pm \Delta_2$	$c \pm \Delta_A$
$\delta^{15}N$	-55.5 ± 17.0	-25.25 ± 55.1	-6.4 ± 9.2
$\delta^{15}N^{SP}$	35.35 ± 6.7	-1.9 ± 11.2	-5.55 ± 5.3
$\delta^{18}O$	23.5 ± 3.0	20.0 ± 6.6	-15.4 ± 20.0

Table 4.2.: Prior distribution parameters taken from Yu, Harris, Lewicka-Szczebak, *et al.* [99] using an additional isotopic measurement in order to increase the estimation accuracy.

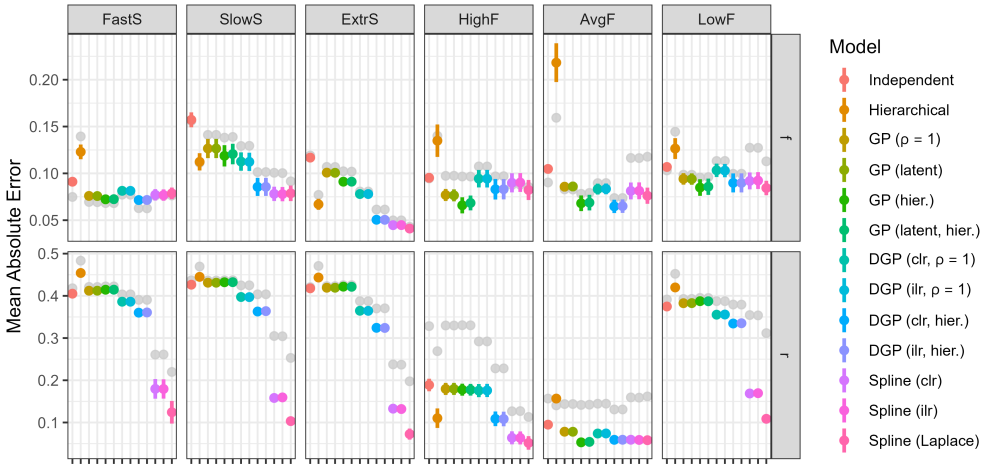

Figure 4.7.: Mean absolute error for all models on $Q = 64$ generated datasets using one additional isotopic measurement which can be found in Table 4.2. The original performance of the models without this additional measurement is shown with gray dots.

Instead of adding additional measurements, more effort could be put into determining the location of the sources and fractionation factor more clearly and thus reduce uncertainty in the input data. To study this case an idealized set of sources and fractionation factor is selected to have mixing and reduction line exactly perpendicular and only having an uncertainty of 10% in each dimension respective to the mean. This renders the mixing and fractionation components independent, since they cannot confound each other.

Measurements sampled in this setting follow exactly the same procedure as in subsection 4.3.1 but only use a Gaussian measurement error with magnitude $\eta = 0.1$. For each example $Q = 64$ datasets were generated and the mean absolute error of estimation is shown in Figure 4.8.

	$b_1 \pm \Delta_1$	$b_2 \pm \Delta_2$	$c \pm \Delta_A$
$\delta^{15}N$	-1 ± 0.2	1 ± 0.2	1 ± 0.1
$\delta^{15}N^{SP}$	1 ± 0.2	-1 ± 0.2	1 ± 0.1

Table 4.3.: Ideal sources and fractionation factor such that mixing and reduction lines are perpendicular with 10% uncertainty in their location.

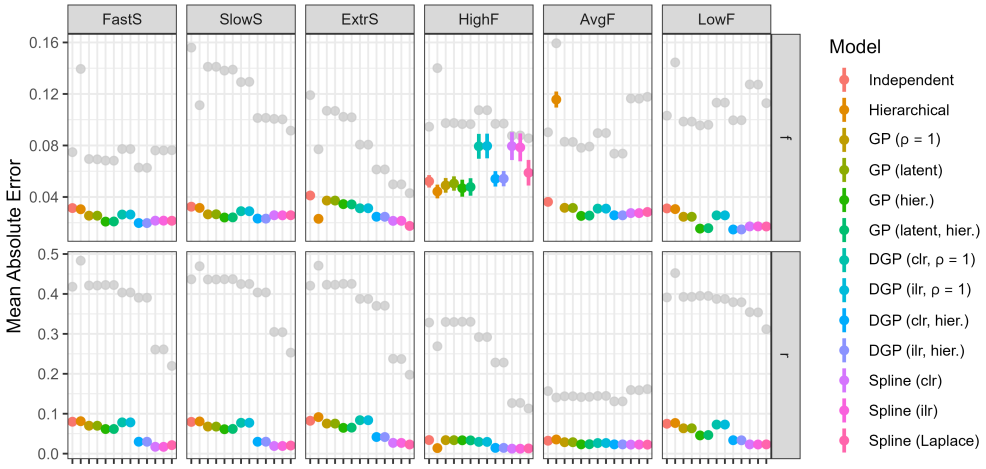

Figure 4.8.: Mean absolute error for all models on $Q = 64$ generated datasets using the idealized sources and fractionation factor with 10% uncertainty in their location found in Table 4.3. The original performance of the models with non-idealized sources is shown with gray dots.

Improving all uncertainties involved to a minimum seems to have great impact on model performance. Almost all mean absolute errors of estimation are below an error margin of 0.05 for source contributions and below 0.1 for fractionation. Furthermore, model choice seems to be less relevant as even the independent time step models perform similarly to the other more sophisticated ones. Interesting is also the fact that the DGP model with fixed hyperparameters as well as the spline GLM with fixed spline basis underperform in source contribution estimation for example HighF. This could be evidence that the default parameters become less robust when noise is removed and they should be selected more carefully.

4.3.3. Influence of Fractionation

Models can have varying performance at different levels of fractionation. For this reason a time series of source contribution values $\mathbf{f}_t^*$ is taken and paired with different constant fractionation values $r_t^* = r^*$ to generate measurements and monitor performance. In total 17 equally spaced fractionation values ranging from $r^* = 0.02$ to $r^* = 0.98$ were used and a total of $Q = 32$ datasets were generated per value. True values for source contributions are taken from the general example GenE used in section 4.4 and a measurement error magnitude of $\eta = 5$ was used.

This experiment is done only with representative models of the four main model classes in order to reduce the number of comparisons that have to be made. The first one being the independent time step model, then Gaussian process prior on measurements with hierarchical estimation of correlation length, DGP prior with hierarchical estimation of correlation length and lastly, B-spline GLM with fixed hyperparameters for degrees of freedom. Mean absolute error of estimation averaged over the dataset simulations and with standard deviations indicated as vertical lines are shown in Figure 4.9.

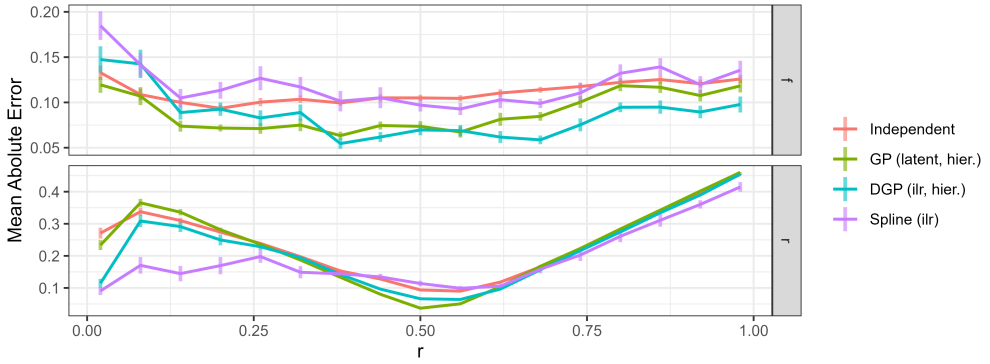

Figure 4.9.: Mean absolute error of the four main model classes over different fractionation weights r. Each reported value is the average over $Q = 32$ dataset simulations with vertical lines indicating standard deviations.

Overall performance shows a clear improvement in the center where $r^* \approx 0.5$. Estimation also becomes more accurate with very low values of r^*, which could mostly be due to the fact that small values have large impacts on measurements and thus estimation can become more accurate. Spline models are expected to perform well here since the time series of fractionation weights is constant which can be reflected by the low degrees of freedom used. The different model classes seem to be equally affected by changes in fractionation otherwise, so presumably the choice of hyperparameters to reflect the situation at hand is more important than selecting a particular model.

Spans of 95% credible intervals can give additional insight into the pattern observed with the estimation accuracy over different values of fractionation weight r^*. If parameter estimation is good then a smaller credible interval span shows a narrow posterior around the correct mean. The width of the 95% credible interval is reported in Figure 4.10.

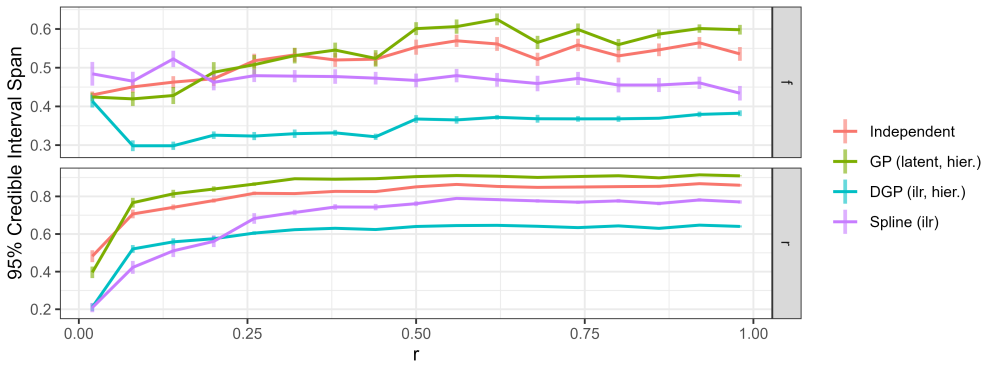

Figure 4.10.: 95% credible interval spans of the four main model classes over different fractionation weights r. Each reported value is the average over $Q = 32$ dataset simulations with vertical lines indicating standard deviations.

DGP prior models have the smallest credible interval span for source contributions and fractionation weights. However, all other models have an interval width of over 0.75 for large fractionation weights and thus span over half of the possible domain. Clearly, due to the Rayleigh fractionation equation being non-linear in r, it is difficult to estimate larger fractionation weights with high accuracy. This could also be an explanation for the dip in estimation error seen in Figure 4.9. If the amount of remaining substrate is larger than 0.1 the data does not give enough information and estimates could group around the prior mean, which is exactly 0.5 and thus gives high estimation accuracy with large credible interval spans.

4.3.4. Influence of Measurement Noise

The main advantage that smooth models such as Gaussian processes and splines have over the independent time step assumption is that they promise to filter measurement noise and thus produce estimates that are more accurate and have a narrower posterior distribution. For this reason an experiment is conducted using values of source contributions $\mathbf{f}_t^*$ and fractionation r_t^* from the general example GenE in section 4.4 to simulate datasets with different measurement noise. Noise values range from $\eta = 0.5$ to $\eta = 20$ and for each separate value a total of $Q = 32$ data sets are generated. Resulting performance of the four main model classes measured as mean absolute error are shown in Figure 4.11.

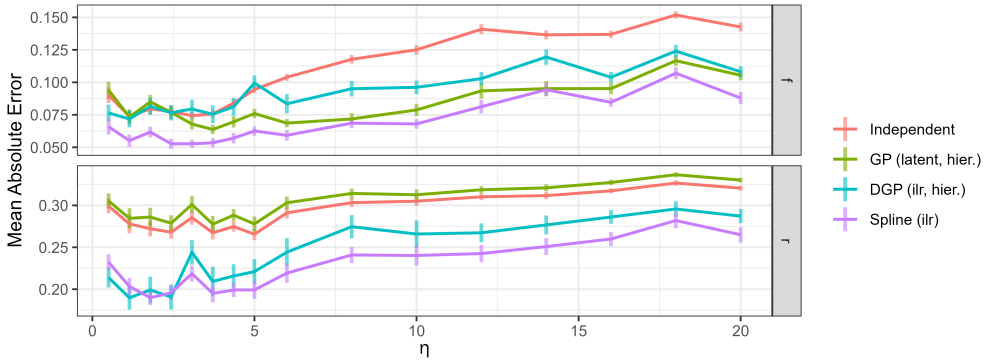

Figure 4.11.: Mean absolute error of estimation of the four main model classes over different measurement noise magnitudes η. Each point represents the average over $Q = 32$ dataset simulations with vertical lines indicating standard deviations.

Performance of all models slowly decreases with increased measurement noise, as is expected. However, it seems like small noise magnitudes do not lead to correspondingly accurate estimates. Especially comparing the results to Figure 4.8 shows that probably most of the estimation error comes from the source variation rather than the measurement noise.

The variance ratio of first differences can be used to assess the quality of the estimated time series in the presence of high frequency changes due to measurement noise. It is shown in Figure 4.12 for the source contribution estimates. Since the rate of change in the fractionation weight is constant for the example chosen it is not suitable for comparison and thus only results regarding the source contributions $\mathbf{f}$ are shown.

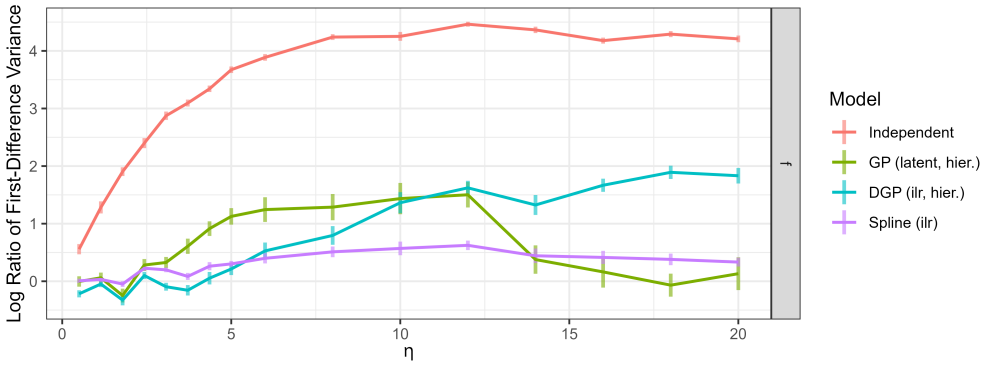

Figure 4.12.: Log variance ratio of first differences of the estimated time series against the true values for source contributions averaged over $Q = 32$ dataset simulations with standard deviations shown as vertical bars.

Variance ratios of the independent time series model gradually increase with increasing measurement noise magnitude. All other models seem to filter the noise well, having much lower overestimations of the first difference variance. The hierarchical DGP model seems to be less equipped to deal with very high noise, which could simply be due to the fact that the weakly informative hyperprior on the correlation length is not suitable here. The spline GLM appears to have constant low values for the ratio of first difference variance possibly due to the fact that the fixed degrees of freedom predetermines the smoothness of the estimates independent of measurement noise.

For each varying level of noise magnitude the sampling time in seconds is measured for the total number of $S = 10000$ samples generated split onto 4 concurrent chains. The reported time in Figure 4.13 is averaged over the $Q = 32$ data simulations with values corresponding to the time it took for the slowest chain to finish.

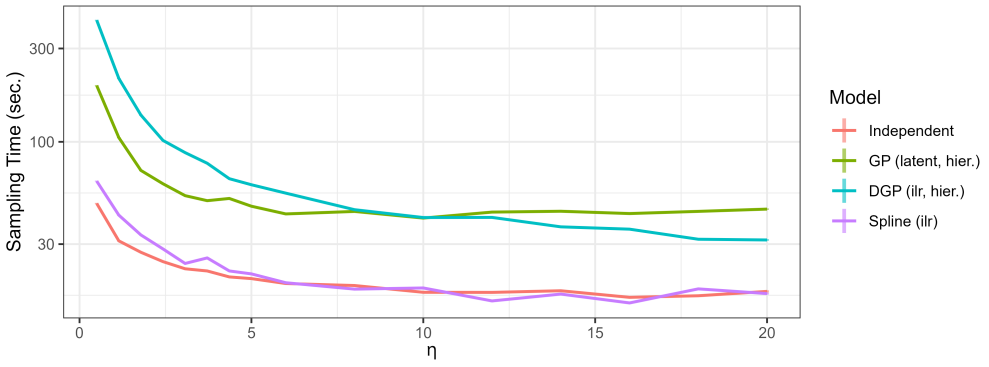

Figure 4.13.: Sampling time in seconds reported for $S = 10000$ posterior samples generated. The values are averaged over $Q = 32$ data simulations and represent the maximum time among 4 concurrent chains.

Fitting times are shifted according to the pattern that was already seen in subsection 4.3.1 with the hierarchical models taking more than double the time. Interesting is the decrease of time needed when errors are large. This might be due to the fact that the model specification takes an estimate of the measurement uncertainty as input and adapts the likelihood accordingly. Likelihoods with larger uncertainty have wider distributions that lead to higher acceptance ratios during Monte Carlo sampling and thus take less time.

4.4. Applications

The main purpose of the time series models is to provide estimates of source contribution and fractionation weights with uncertainty. In the sections above, only the performance metrics aggregated over many simulations have been shown. To illustrate the modeling capabilities a representative general example (GenE) is simulated from fixed parameter values and the inference results are shown in comparison to the true values. Real datasets are available as well, but interpretation of results without having ground truths available is generally hard and thus not particularly useful. Nevertheless, real data results are available in Appendix C.

Source isotopic signatures as well as fractionation factor priors are taken from Yu, Harris, Lewicka-Szczebak, *et al.* [99] in Table 3.1, with one single value of $\mathbf{S}$ and A sampled to generate the dataset. The mixing equation using Rayleigh fractionation is used with the ground truth parameters to compute measurement means and a Gaussian error with magnitude $\eta = 5$ is used to sample $N = 64$ measurements $X_1, \ldots, X_N$. The fixed parameter values and the simulated data is shown in Figure 4.14.

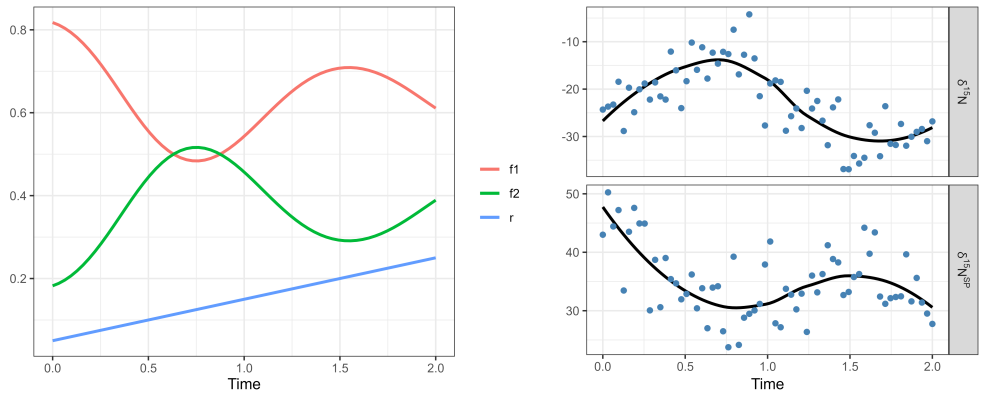

Figure 4.14.: Fixed ground truth values of source contributions and fractionation used to simulate measurements in the left panel and the measurement values simulated accordingly together with LOESS estimates in the right panel.

Normally, the dataset would need to be preprocessed as indicated by the smoother in Figure 4.14. Tools such as dual isotope plots for visualization and smoothing methods such as local linear regression are readily available. However, the generated dataset is known to have feasible values for measurements, because it was generated that way and the measurement error of $\eta = 5$ used for the simulation is small enough that no preprocessing is needed.

In order to run the Bayesian models and estimate source contributions and fractionation over time the auxiliary distributions of the source isotopic signatures $\mathbf{S}$ and the fractionation factor A as well as the noise magnitude η must be supplied in addition to the dataset. Three different model classes are run to illustrate the computed output. The first one is the independent time step model described in Equation 3.5, then the spline GLM described in Equation 3.11 and lastly the hierarchical DGP prior model described in Equation 3.10.

From the output that the models produce either summary statistics of the posterior such as its mean and its quantiles can be gathered or the full sample trace of all posterior time series sampled can be extracted. The estimated means are shown in Figure 4.15 together with the shaded regions representing the 95% equally-tailed credible interval.

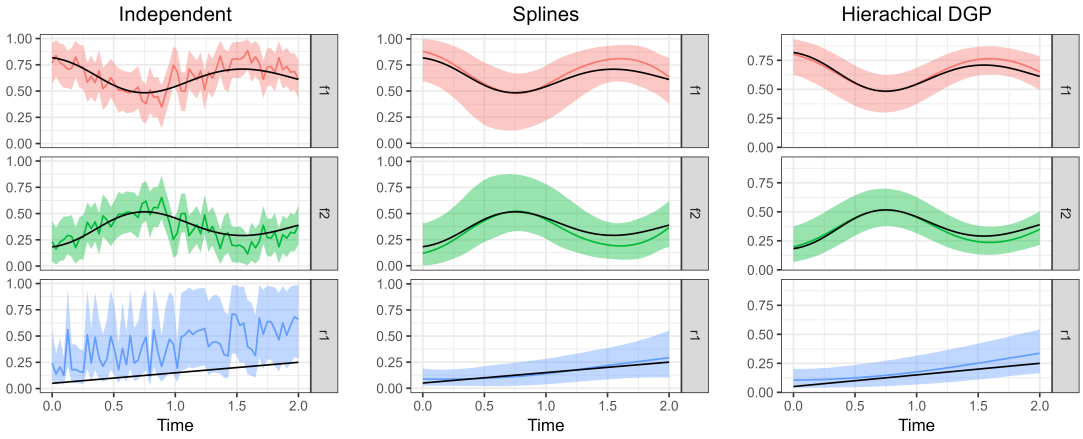

Figure 4.15.: Posterior means of the three model types compared to the true parameter values. Shaded areas indicate 95% credible intervals and the true parameter values used to simulate the measurements are shown as black lines.

The independent time step model clearly shows that the measurement error has a large effect on the estimated parameters. Nevertheless, the credible interval covers the true parameter values well and is reasonably narrow. Fractionation weight estimaton seems to be biased toward higher values which could be due to confounding with the variation in source isotopic signatures. The B-spline basis for the GLM seems to have default values for degrees of freedom that are fairly optimal in this case. The time series of parameter estimates is now similarly smooth to the actual parameter series. Estimation using the hierarchical DGP prior model seems to give the best results visually. The time series are adequately smooth and estimates are close to the true values with narrow credible intervals.

The posterior distribution of the models can be fully extracted in which case it can be interpreted as a distribution over time series functions. As such not only summary statistics can be plotted but the whole distribution as well. Similarly to plotting a dataset consisting of individual points, certain techniques exist to visualize the shape of the distribution. Using functional data analysis the functions can be associated with band depth thus ordering them from the most central time series to the most extremal ones. The distribution can then be plotted using functional boxplots [106], [107], which display the median time series as a black line and a band of the inner most

50% of the curves as shaded region shown in Figure 4.16. Additional lines act as boundaries to conjectured outlier curves similar to whiskers in regular boxplots and curves contained in the sample that are considered outliers are separately plotted using dotted lines.

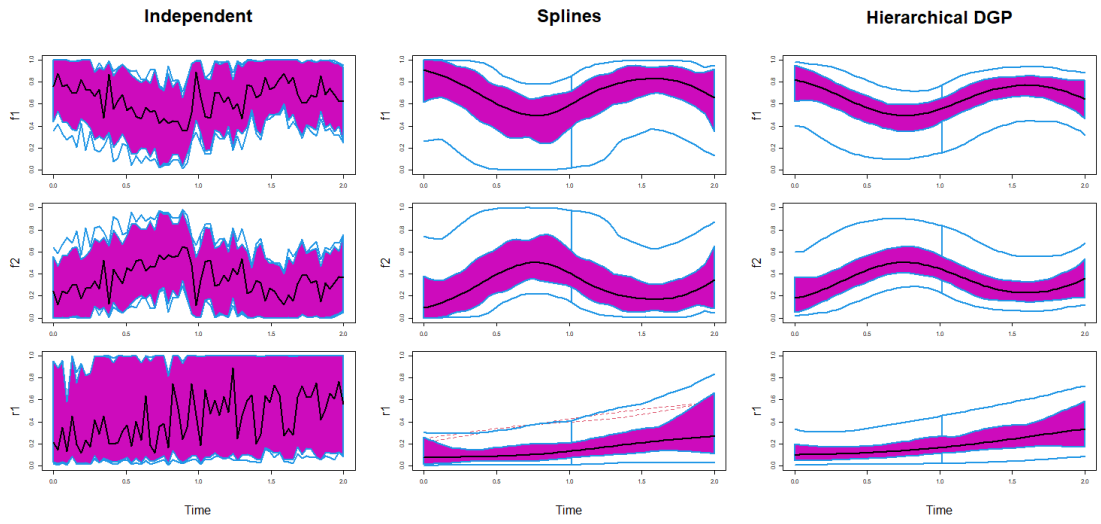

Figure 4.16.: Functional boxplots of the three model types for variables f_1, f_2 and r from top row to bottom row. Visualizations are computed using the *Functional Data Analysis* package in R [108].

The independent time step model covers a wide range of the domain with the sampled curves which is mainly because it does not take time series information into account that would reduce uncertainty. The B-spline GLM has much narrower functional boxes with small bumps indicating the location of spline knots that were used to construct and evaluate the basis. The hierarchical DGP model has the smallest functional box, which seems to be as smooth as the time series of true parameter values as well, thus indicating that not only the summary statistics have favorable properties, but also the individual samples generated.

4.4.1. Selecting Fractionation Equations

Posterior distribution samples can be used for many further applications going beyond simply examining its summary statistics. The focus for the experiments above was on closed systems and the Rayleigh fractionation equation having $\mu_{\text{closed}}(\mathbf{F}, \mathbf{r}) = \mathbf{SF} + A \log \mathbf{r}$ to account for fractionation in addition to mixing. However, in practice, it is oftentimes unclear if this is an adequate representation of the system with an alternative for nitrous oxide reduction being open-system fractionation $\mu_{\text{open}}(\mathbf{F}, \mathbf{r}) = \mathbf{SF} - A(\mathbf{1} - \mathbf{r})$.

Resolving this question essentially requires the comparison of two likelihood formulations $p_{\text{closed}}(\mathbf{X}|\mathbf{F}, \mathbf{r})$ and $p_{\text{open}}(\mathbf{X}|\mathbf{F}, \mathbf{r})$, since the mixing and fractionation equation computes measurement means of a Gaussian distribution that the measurements are subject to. Traditionally, in frequentist statistics, likelihood-ratio tests have been used in conjunction with the maximum

likelihood estimates of the parameters to select the best option. In Bayesian statistics an analogous device are Bayes factors, which compare model formulations in terms of their marginal likelihood instead.

$$p(\mathbf{X}) = \int p(\mathbf{X}|\mathbf{F}, \mathbf{r})\pi(\mathbf{F})\pi(\mathbf{r})d\mathbf{F}d\mathbf{r} \tag{4.1}$$

Evaluating the marginal likelihood, also called evidence, is not trivial even if posterior samples of the parameters exist. However, resampling techniques such as bridge sampling can be employed to accomplish precisely this [109]. The Bayes factor of closed-system fractionation over open-system fractionation is then the the ratio of marginal likelihoods.

$$\mathrm{BF}_{\mathrm{closed}} = \frac{p_{\mathrm{closed}}(\mathbf{X})}{p_{\mathrm{open}}(\mathbf{X})} = \frac{\int p_{\mathrm{closed}}(\mathbf{X}|\mathbf{F}, \mathbf{r})\pi(\mathbf{F})\pi(\mathbf{r})d\mathbf{F}d\mathbf{r}}{\int p_{\mathrm{open}}(\mathbf{X}|\mathbf{F}, \mathbf{r})\pi(\mathbf{F})\pi(\mathbf{r})d\mathbf{F}d\mathbf{r}} \tag{4.2}$$

Resulting values for the Bayes factor can be understood as evidence for or against a given hypothesis and interpretations of values in certain ranges are tabulated [45]. For this simulated example the dataset was generated using closed-system fractionation and consequently the computed Bayes factor has $\mathrm{BF} > 10^5$ which counts as decisive evidence for the description of the data using closed-system fractionation.

Care must be taken when comparing likelihoods with Bayes factors since normalized prior distributions are not needed for estimating the posterior and thus not all models satisfy this. If the source isotopic signatures $\mathbf{S}$ use a mixture of uniform spread and Gaussian error then their prior will be unnormalized and bridge sampling will no longer give correct estimates of the marginal likelihood. Furthermore, the Jeffreys prior for the fractionation index is improper and thus cannot be normalized and used for Bayes factor computation.

4.5. Limitations

Posterior predictive checks are a useful tool for Bayesian model comparison. They take full posterior distributions into account and are applicable for real datasets where true parameter values are not known beforehand. The models were constructed and compared with the setting in mind that only one singe time series would be available when the models are used. This makes the estimation of expected predictive densities impossible, since in essence the single time series represents only one single data point. In practice, one must thus rely on pointwise predictive densities evaluated at this single time series instead or resort to Bayes factors or other comparison methods. Simulations of cases where multiple time series would be accessible were not conducted and the model implementations do not account for it.

Spline GLMs outperform Gaussian process based models in the estimation of fractionation weights close to 0 or 1. This could be due to the fact that the spline bases include intercept terms which allows for extremal fractionation values more easily. Since for source contributions such a term is not needed, it was omitted from Gaussian process models entirely and causes their estimates for fractionation weights to be biased. Including it could serve as a means to improve them even more.

Extension of the presented models to multiple continuous dimensions for instance required for spatio-temporal modeling would be natural for Gaussian processes as well as splines and in subsection 2.4.3 an alternative approach is suggested. Future datasets could make great use for such modeling capabilities, but no experiments or implementations could be provided due to scope constraints.

The mixing and fractionation equations that are used such open-system fractionation equation use the assumption of stationarity of fluxes for their derivation. Using these equations for time series models that are implicitly non-stationary could have unintended consequences for their validity. A good further step could therefore be to examine the derivation of fractionation equations via differential equations closely, while taking non-stationarity into account.

Simulations used for model comparison tend to have only few repetitions which makes definitive statements about model performance much more difficult. The behavior of most models could still be assessed, but especially for spline GLMs it remains unclear whether the chosen defaults are truly as reliable as they seem and can perform in a general context.

Chapter 5

Conclusion

5.1. Summary

This thesis was conducted with the goal of studying Bayesian inference techniques for the analysis of isotopic data and develop approaches that are suitable for cases where mixing and fractionation is present as well as where time series information is available. To accomplish this, a few related methods of estimating source contributions were surveyed, with the most common one being isotopic mapping that can be reduced to solving linear systems of equations and hence can estimate uncertainty only under restrictive assumptions. Most notable Bayesian techniques include 3DIM [17] which extends previous formulations to more than two isotopic measurements, FRAME [18] which implements Bayesian inference for stationary cases and Bayesian mixing model implementations [20]–[22] typically used for animal diet partitioning.

A total of four time series model classes were presented in section 3.2 which are independent time step models, Gaussian process priors on measurements, Dirichlet-Gaussian process priors and generalized linear models with spline bases. All classes were implemented in different variations using Stan and thoroughly compared in various scenarios. They were constructed for the purpose of finding source contributions from isotopic measurements of nitrous oxide and are able to incorporate fractionation equations, making them a direct extension to FRAME.

Model comparison in section 4.3 showed that, while it seemed like there are use cases for all classes, hierarchical Dirichlet-Gaussian process models tended to be best in terms of estimation accuracy, narrowness of posterior and matching of time series properties such as curvature. However, the joint estimation of correlation length parameters lead to high fitting times in the range of 60 seconds for time series of 32 points. Generalized linear models with spline bases were fast and similarly accurate, but do not explicitly tune hyperparameters. Nevertheless, the defaults used seem to be reliable and outperform all other model classes in certain cases where the effect of fractionation weight is extremal and slowly changes in time.

Experiments on improvements of the data concluded that the most effective strategy to refine source contribution and fractionation estimates seems to be to reduce uncertainty in the input data with noise magnitudes of less than 10% giving very accurate estimates. Rendering the system of equations overdetermined by adding additional measurements could be fruitful as well, although care has to be taken that they are sufficiently informative and do not introduce more noise. Inference on consumption levels was underwhelming, especially if the amount of remaining substrate was either very high or very low. Alternative priors for the fractionation weight showed little difference between multiple choices, but they could confirm that the uniform prior, that was already in use before, performs best on average for inference of source contributions. Finally,

different levels of measurement error had rather similar performance, pointing to the fact that a substantial amount of estimation error can potentially be linked to uncertainty of the input data, namely source isotopic signatures and fractionation factors.

The presented models were implemented together with some of the original isotopic data analysis techniques in the `TimeFRAME` package. The implementation is in `R` using Stan as sampler and its description as well as the available functions are referenced in Appendix A.

5.2. Recommendations

Given the model comparisons made in section 4.3 it seems like most models perform well in most cases. It is therefore fine to use independent time step assumptions in cases where only pointwise estimates are needed. The hierarchical formulation of Dirichlet-Gaussian process models are a high performing alternative that seems to be more accurate, have narrower posterior distributions and incorporate time series information well. They look to be a good choice for a wide range of applications if the longer fitting times are able to be managed. Generalized linear models with spline bases have similar high performance, but much shorter fitting times. However, their application requires the modeler to construct a spline basis where many different choices are possible. Default values are provided and seem to be robust according to the experiments conducted, but it is unclear whether this generalizes to real datasets.

When using Gaussian processes the choice of long correlation lengths can lead to long fitting times due to the estimation of many highly correlated variables. If this is desired nonetheless, it is useful to aggregate or subsample the data beforehand. It remains unclear what hyperparameters such as correlation length or spline degrees of freedom work best in practice and so the chosen default values are not necessarily suitable or even valid in different scenarios. Hierarchical models can help with estimating hyperparameters from the data, so a sensible course of action could be to use wide hyperpriors and use posterior mode estimates for data of similar character.

Applying Bayesian models to real data can be challenging if the measurement noise is overwhelming and the measured values are out of proportion. In addition to giving estimates with high uncertainty, the fitting time can become exceptionally long if the measurements are not close to any solution that can be produced with mixing and fractionation. Much care should therefore be taken to clean and preprocess input data properly and to set appropriate values for the measurement noise parameter when fitting the models.

5.3. Future Work

The generalized linear model has a multitude of extension possibilities and variations that are not sufficiently examined here. Similar to Stock, Jackson, Ward, *et al.* [20] mixed effects can be included for measurement repetitions together with the possibility of adding other covariates corresponding to exogenous variables. Furthermore, spline bases are highly flexible in the way they are constructed and hence allow for alternative choices such as the type of spline, the degrees of freedom and the placement of knots. These variations could be studied in-depth and optimized for best performance and desired properties.

The models presented in this thesis provide groundwork on how to deal with fractionation, time series information and uncertainties simultaneously. Most datasets available at the current time suffer from high uncertainty in the input data and thus conclusions are still vague. Future developments in the direction of refining the estimates on source isotopic signatures and additional measurement capabilities could lead to higher quality of analyses. Measurements of clumped N_2O isotopes are currently being studied and could open up the possibility of using up to 7-dimensional input data [110].

Interesting approaches where multiple smoothing dimensions are present are measurements that are taken with respect to space and time. Simultaneous spatio-temporal estimation could be approached by extending the Gaussian process priors and spline bases to multiple dimensions with the latter extension being as simple as adding spline expansions in each dimension. Both additions are essentially possible with the framework presented in this thesis, but still need extensive testing and experimentation. Furthermore, formulating the problem as an inhomogeneous differential equation could lead to entirely different estimation methods involving partial differential equation solvers that are currently being developed [81].

Appendix A

R Package

A.1. Description

`TimeFRAME` is a data analysis package that can be used for Bayesian estimation of source contributions and fractionation of isotopic measurement time series. It uses Bayesian parameter estimation with Stan to estimate uncertainty and produce posterior samples. Additionally, the package provides utility functions to sample from prior distributions and simulate measurements.

A.2. Methods

`frame_model` Creates a model specification of class `FrameModel` for running statistical samplers on a dataset of isotopic measurements.

- `sources` Data matrix with $2d$ columns corresponding to d isotopic measurements and d spreads and K rows representing sources.
- `frac` Data matrix with $2d$ columns corresponding to d isotopic measurements and d spreads and L rows representing fractionation factors (Default empty `data.frame`).
- `x` Data matrix with d columns corresponding to isotopic measurements and N rows representing measurement repetitions (Default empty `data.frame`).
- `sd` Data matrix with d columns corresponding to standard deviations and N rows representing repetitions (Default 0).
- `t` Vector with N elements corresponding to time points or alternative continuous orderings of measurements (Default `NULL`).

`isotope_map` Computes source contribution and fractionation weight by solving linear systems of equations using the isotope mapping approach.

- `model` Object of class `FrameModel`.
- `x` Data matrix with d columns corresponding to isotopic measurements and N rows representing repetitions. Overrides parameters of `model`.
- `frac.eq` Type of fractionation equation to use. Must be one of `"open|closed"` (Default `"closed"`).

`fit_stationary` Fits a Stan model using original FRAME formulation treating measurements as independent repetitions. Ignores time series information and fits one set of parameters.

`model` Object of class `FrameModel`.

`x` Data matrix with d columns corresponding to isotopic measurements and N rows representing repetitions. Overrides parameters of `model`.

`eta` Numeric, indicating additional measurement noise (Default 0).

`fit_frame` Fits a Stan model using original frame formulation treating measurements and estimates as independent in time.

`model` Object of class `FrameModel`.

`x` Data matrix with d columns corresponding to isotopic measurements and N rows representing repetitions. Overrides data matrix of `model`.

`sd` Data matrix with d columns corresponding to standard deviations and N rows representing repetitions (Default 0).

`eta` Numeric, indicating additional measurement noise (Default 0).

`fit_gp` Fits a Stan model using a GP model (Gaussian Process Prior on Measurements).

`model` Object of class `FrameModel`.

`x` Data matrix with d columns corresponding to isotopic measurements and N rows representing repetitions. Overrides data matrix of `model`.

`sd` Data matrix with d columns corresponding to standard deviations and N rows representing repetitions (Default empty 0).

`t` Vector with N elements corresponding to time points or alternative continuous orderings of measurements (Default `NULL`).

`rho` Numeric indicating scaled correlation length of measurements (Default 1).

`estim.rho` Logical indicating whether correlation length should be estimated by a hierarchical model (Default `FALSE`).

`eta` Numeric, indicating additional measurement noise (Default 0).

`fit_dgp` Fits a Stan model using a DGP model (Dirichlet-Gaussian Process Prior).

`model` Object of class `FrameModel`.

`x` Data matrix with d columns corresponding to isotopic measurements and N rows representing repetitions. Overrides data matrix of `model`.

`sd` Data matrix with d columns corresponding to standard deviations and N rows representing repetitions (Default 0).

`t` Vector with N elements corresponding to time points or alternative continuous orderings of measurements (Default `NULL`).

`sigma` Numeric indicating concentration of the Dirichlet priors (Default 1).

`rho` Numeric indicating scaled correlation length of source contributions (Default 1).

`rho.r` Numeric indicating scaled correlation length of fractionation (Default 1).

`estim.rho` Logical indicating whether correlation lengths should be estimated by a hierarchical model (Default `FALSE`).

`eta` Numeric, indicating additional measurement noise (Default 0).

`fit_glm` Fits a Stan model using a Spline GLM model (Generalized Linear Model with Spline Basis).

`model` Object of class `FrameModel`.

`x` Data matrix with d columns corresponding to isotopic measurements and N rows representing repetitions. Overrides data matrix of `model`.

`sd` Data matrix with d columns corresponding to standard deviations and N rows representing repetitions (Default 0).

`t` Vector with N elements corresponding to time points or alternative continuous orderings of measurements (Default `NULL`).

`M` Numeric indicating spline degrees of freedom for source contributions (Default 8).

`M.r` Numeric spline degrees of freedom for fractionation (Default 4).

`eta` Numeric, indicating additional measurement noise (Default 0).

`sample_dgp` Creates samples from the prior distribution of the DGP model. Used to check if hyperparameters produce the expected output.

`n` Number of samples to be generated.

`t` Vector of N time points where the sampled functions should be evaluated.

`rho` Numeric indicating scaled correlation length of the functions to be sampled.

`sigma` Numeric indicating concentration of the functions to be sampled.

`sample_spline` Creates samples from the prior distribution of the Spline GLM model. Used to check if hyperparameters produce the expected output.

`n` Number of samples to be generated.

`t` Vector of N time points where the sampled functions should be evaluated.

`M` Integer indicating number of degrees of freedom of the functions to be sampled.

`sigma` Numeric indicating concentration of the functions to be sampled.

`sample_means` Creates samples for measurement means with given source contribution and fractionation weights.

`model` Object of class `FrameModel`.

`f` Data frame with $K + L$ columns corresponding to values of source contribution and fractionation weights and N rows corresponding to different time points.

`sample_measurements` Creates samples from the prior distributions that are used for time series modeling. Use this to check if hyperparameters produce what you expect.

`model` Object of class `FrameModel`.

`f` Data frame with $K + L$ columns corresponding to values of source contribution and fractionation weights and N rows corresponding to different time points.

`eta` Numeric, indicating measurement noise to be used or data frame of measurement noises with d columns corresponding to isotopic measurements and N rows corresponding to time points (Default 1).

Appendix B

Experiments

All experiments were conducted by fixing true values of source contribution time series $\mathbf{F}^* \in \mathbb{R}^{K \times N}$ and fractionation weights $\mathbf{r}^* \in \mathbb{R}^{1 \times N}$ and simulating Q measurements $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(Q)} \in \mathbb{R}^{d \times N}$ from them by using the likelihood $p(\mathbf{X}|\mathbf{F}^*, \mathbf{r}^*)$ derived from the mixing equation and the measurement distribution. Model fitting yielded S posterior samples for source contributions $\mathbf{F}^{(q,1)}, \ldots, \mathbf{F}^{(q,S)}$ and fractionation weights $\mathbf{r}^{(q,1)}, \ldots, \mathbf{r}^{(q,S)}$ for runs $q = 1, \ldots, Q$. From these samples pointwise estimators corresponding to posterior means $\hat{\mathbf{F}}^{(q)} = \frac{1}{S} \sum_{s=1}^{S} \mathbf{F}^{(q,s)}$, $\hat{\mathbf{r}}^{(q)} = \frac{1}{S} \sum_{s=1}^{S} \mathbf{r}^{(q,s)}$ and medians $\tilde{\mathbf{F}}^{(q)} = \mathbf{F}^{(q,[\frac{S}{2}])}$, $\tilde{\mathbf{r}}^{(q)} = \mathbf{r}^{(q,[\frac{S}{2}])}$ as well as posterior quantiles $q(\gamma|\mathbf{X}^{(q)})$ could be computed.

The following metrics were computed per run $q = 1, \ldots, Q$ and aggregated to compare model performance. Metrics for $\mathbf{r}$ are computed analogously:

- Root Mean Squared Error of Posterior Mean (RMSE)

$$\sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(\hat{\mathbf{F}}_{kt}^{(q)} - \mathbf{F}_{kt}^* \right)^2}$$

- Mean Absolute Error of Posterior Mean (MAE)

$$\frac{1}{N} \sum_{t=1}^{N} \left| \hat{\mathbf{F}}_{kt}^{(q)} - \mathbf{F}_{kt}^* \right|$$

- Log Mean Percentage Error of Posterior Mean (LMAPE)

$$\frac{1}{N} \sum_{t=1}^{N} \log \left| \frac{\hat{\mathbf{F}}_{kt}^{(q)} - \mathbf{F}_{kt}^*}{\mathbf{F}_{kt}^*} \right|$$

- Log Variance Ratio of First Differences of Posterior Mean (RDVar)

$$\log \frac{\sum_{t=1}^{N-1} \left(\Delta \mathbf{F}_{kt}^{(q)} - \frac{1}{N} \sum_{j=1}^{N-1} \Delta \mathbf{F}_{kj}^{(q)} \right)^2}{\sum_{t=1}^{N-1} \left(\Delta \mathbf{F}_{kt}^* - \frac{1}{N} \sum_{j=1}^{N-1} \Delta \mathbf{F}_{kj}^* \right)^2}$$

- Log Correlation Ratio of First Differences of Posterior Mean (RDCor)

$$\log \frac{\sum_{t=1}^{N-1} \rho(\mathbf{F}_{k}^{(q)}, \mathbf{F}_{k,t+1}^{(q)})^2}{\sum_{t=1}^{N-1} \rho(\mathbf{F}_{kt}^{*}, \mathbf{F}_{k,t+1}^{*})^2}$$

- Root Mean Squared Error of Posterior Median (RMSE_med)
- Mean Absolute Error of Posterior Median (MAE_med)
- Log Mean Percentage Error of Posterior Median (LMAPE_med)
- Log Variance Ratio of First Differences of Posterior Median (RDVar_med)
- Log Correlation Ratio of First Differences of Posterior Median (RDCor_med)
- 50% Credible Interval Coverage (Cov50)

$$\frac{1}{N} \sum_{t=1}^{N} \mathbb{1} \left\{ q_t(0.25|\mathbf{X}^{(q)}) \leq \mathbf{F}_{kt}^{*} \leq q_t(0.75|\mathbf{X}^{(q)}) \right\}$$

- 95% Credible Interval Coverage (Cov95)

$$\frac{1}{N} \sum_{t=1}^{N} \mathbb{1} \left\{ q_t(0.025|\mathbf{X}^{(q)}) \leq \mathbf{F}_{kt}^{*} \leq q_t(0.975|\mathbf{X}^{(q)}) \right\}$$

- Posterior Standard Deviation (SD)

$$\frac{1}{N} \sum_{t=1}^{N} \sqrt{\frac{1}{S-1} \sum_{s=1}^{S} \left(\mathbf{F}_{kt}^{(q,s)} - \hat{\mathbf{F}}_{kt}^{(q)} \right)^2}$$

- 95% Credible Interval Span (Span)

$$\frac{1}{N} \sum_{t=1}^{N} q_t(0.975|\mathbf{X}^{(q)}) - q_t(0.025|\mathbf{X}^{(q)})$$

- Posterior Tail Probability (PVal)

$$\frac{2}{N} \sum_{t=1}^{N} \min \left\{ \frac{1}{S} \sum_{s=1}^{S} \mathbb{1} \left\{ \mathbf{F}_{kt}^{(q,s)} \geq \mathbf{F}_{kt}^{*} \right\}, \frac{1}{S} \sum_{s=1}^{S} \mathbb{1} \left\{ \mathbf{F}_{kt}^{(q,s)} \leq \mathbf{F}_{kt}^{*} \right\} \right\}$$

- Posterior Mean of RMSE (rmse_mean)

$$\frac{1}{S}\sum_{s=1}^{S}\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\hat{\mathbf{F}}_{kt}^{(q,s)}-\mathbf{F}_{kt}^{*}\right)^{2}}$$

- Posterior Standard Deviation of RMSE (rmse_sd)
- Posterior Mean of MAE (mae_mean)

$$\frac{1}{S}\sum_{s=1}^{S}\frac{1}{N}\sum_{t=1}^{N}\left|\hat{\mathbf{F}}_{kt}^{(q,s)}-\mathbf{F}_{kt}^{*}\right|$$

- Posterior Standard Deviation of MAE (mae_sd)
- Runtime in Seconds (Runtime)
- Log Pointwise Predictive Density (LPD)

$$\log\frac{1}{S}\sum_{s=1}^{S}p(\mathbf{X}^{(q)}|\mathbf{F}^{(q,s)})$$

- Expected Log Pointwise Predictive Density by WAIC (ELPD) [102], [111]

Figure B.1.: (a) Illustrative Examples (subsection 4.3.1).

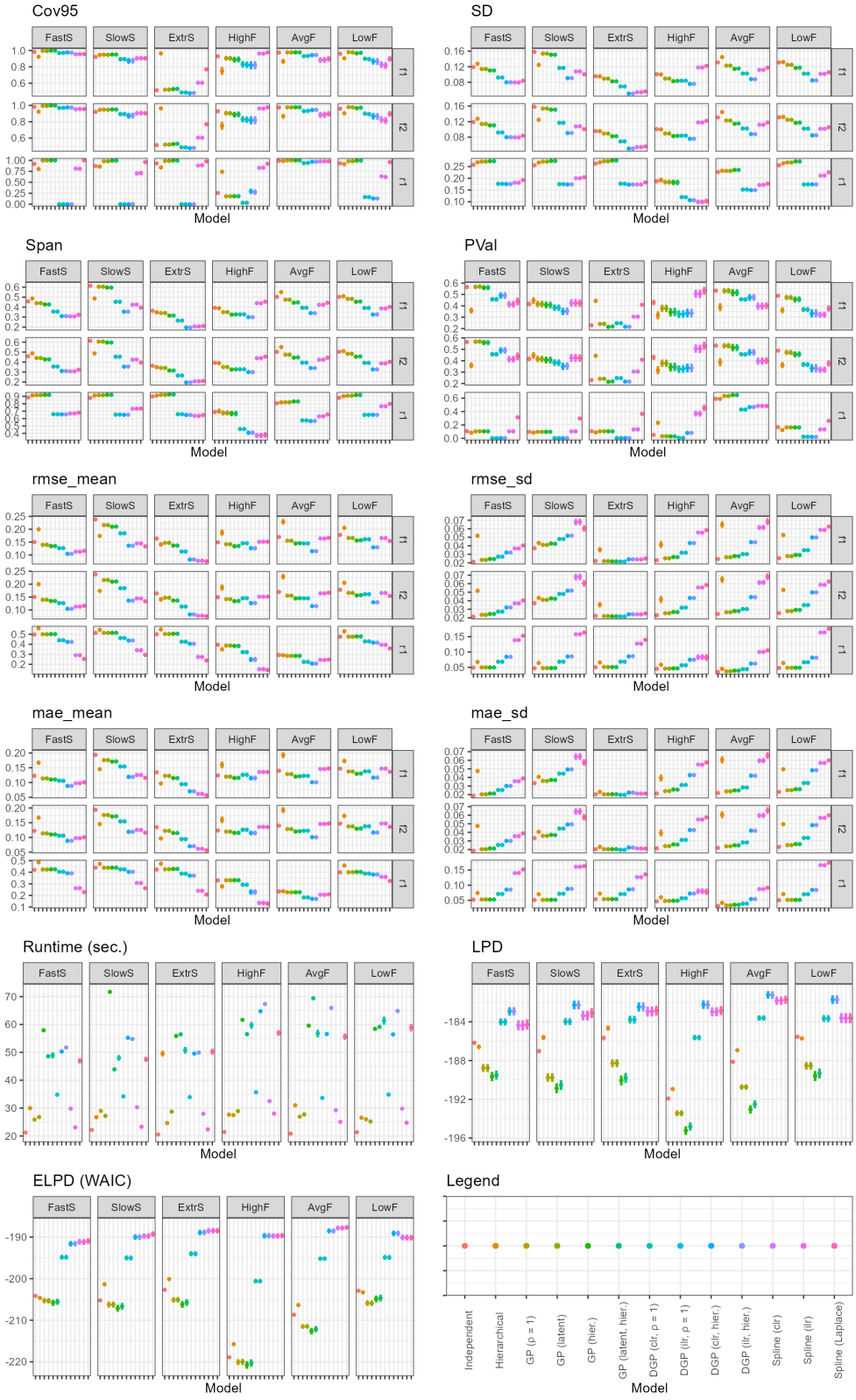

Figure B.2.: (b) Illustrative Examples (subsection 4.3.1).

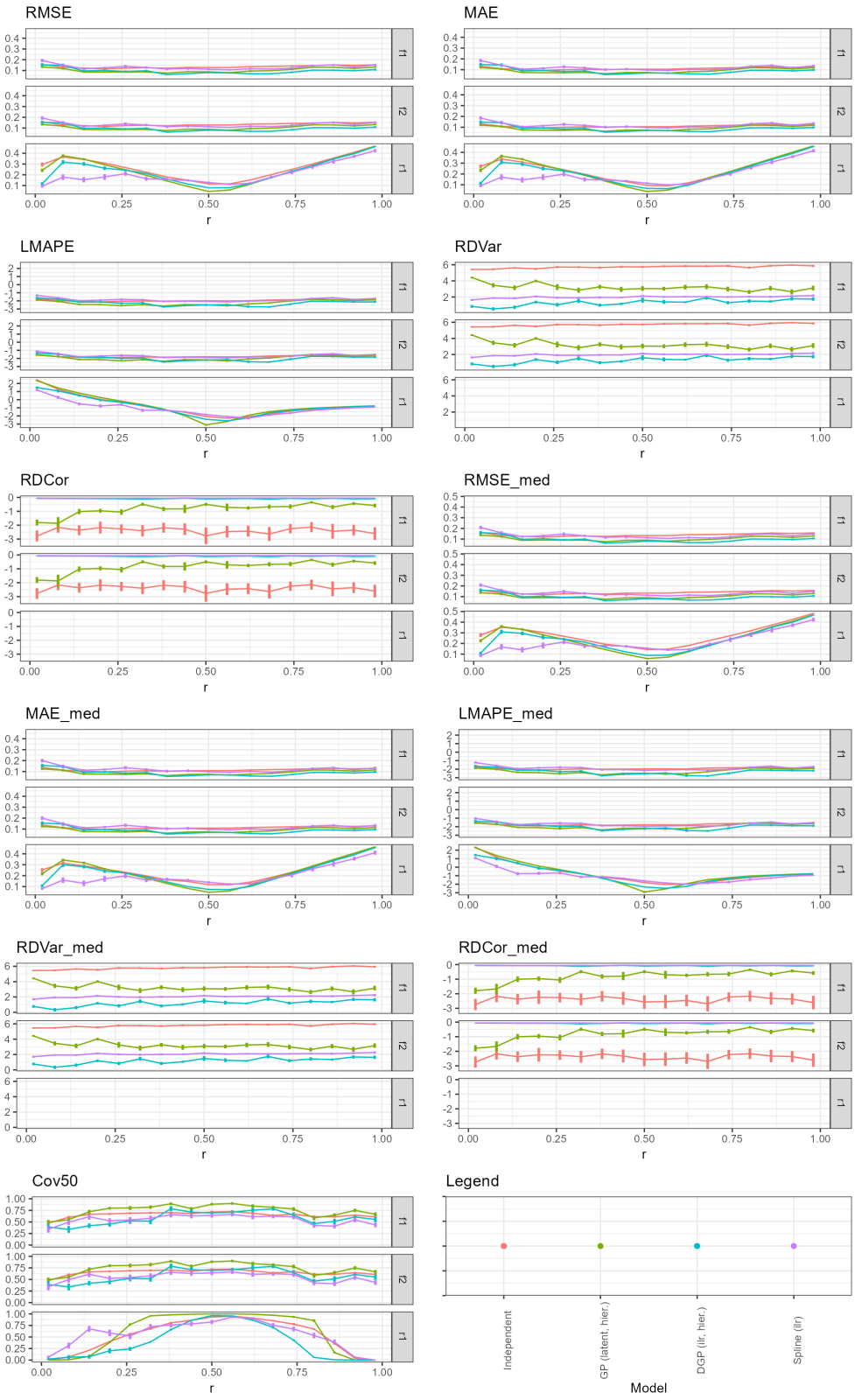

Figure B.3.: (a) Influence of Fractionation (subsection 4.3.3).

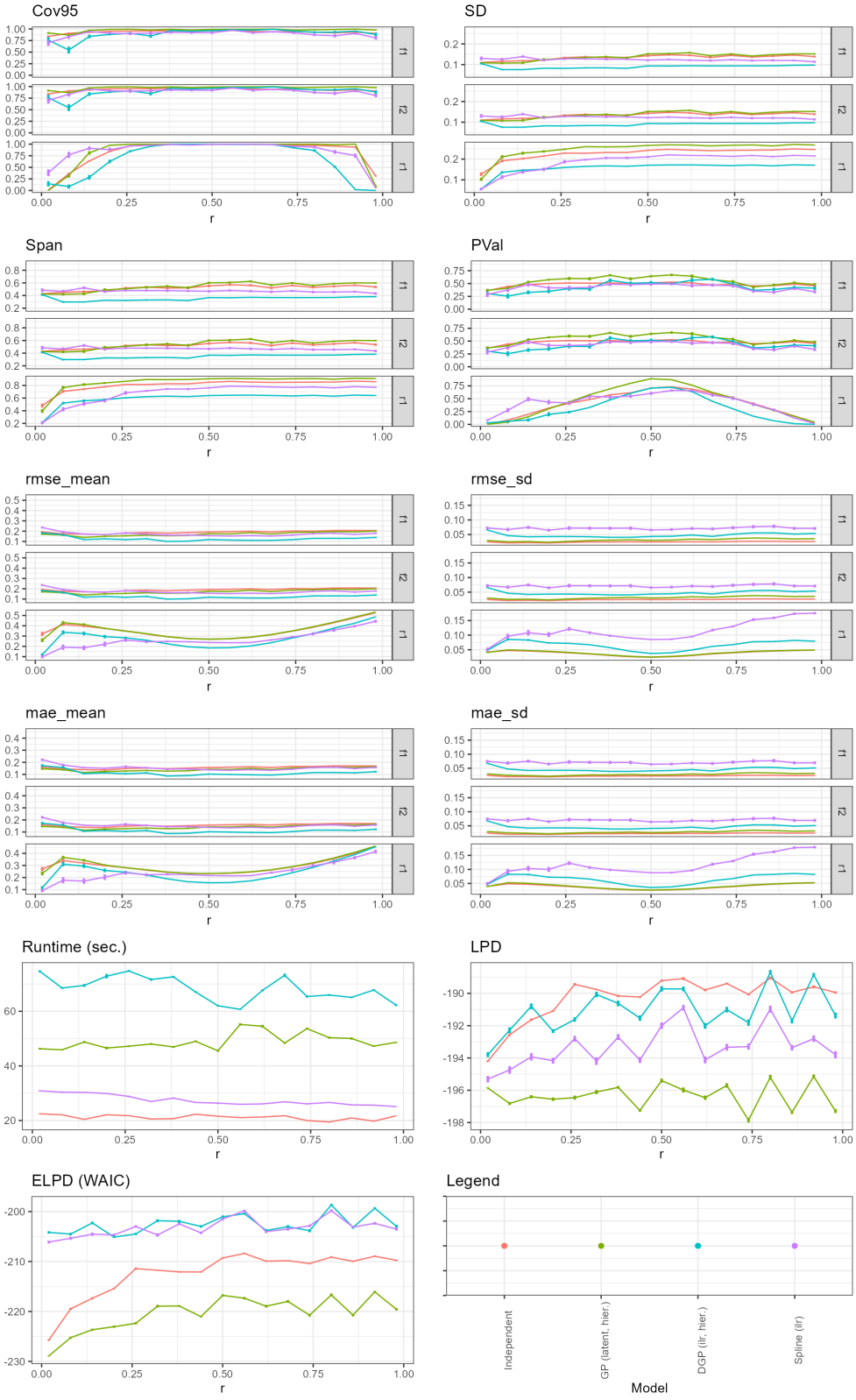

Figure B.4.: (b) Influence of Fractionation (subsection 4.3.3).

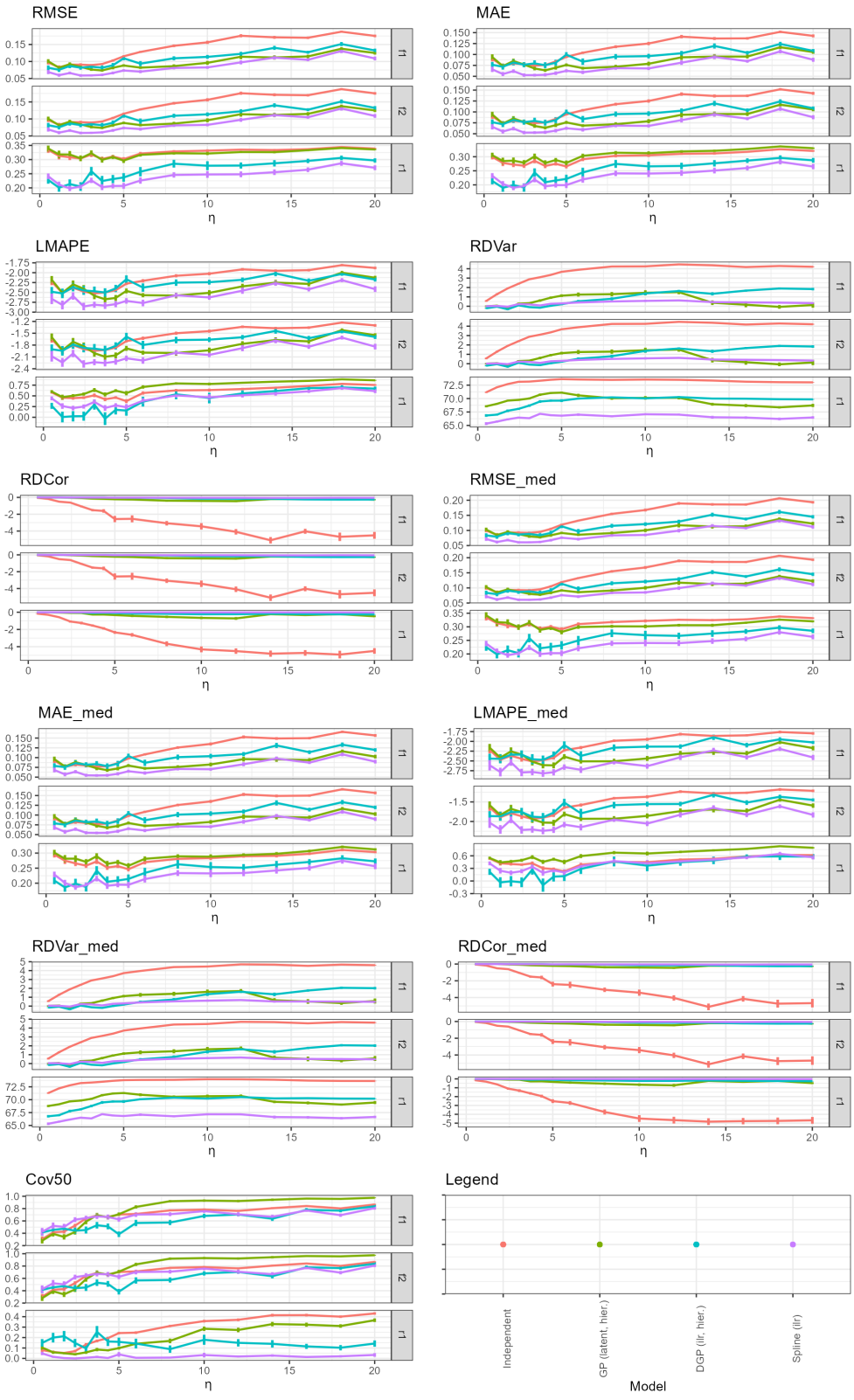

Figure B.5.: (a) Influence of Measurement Noise (subsection 4.3.4).

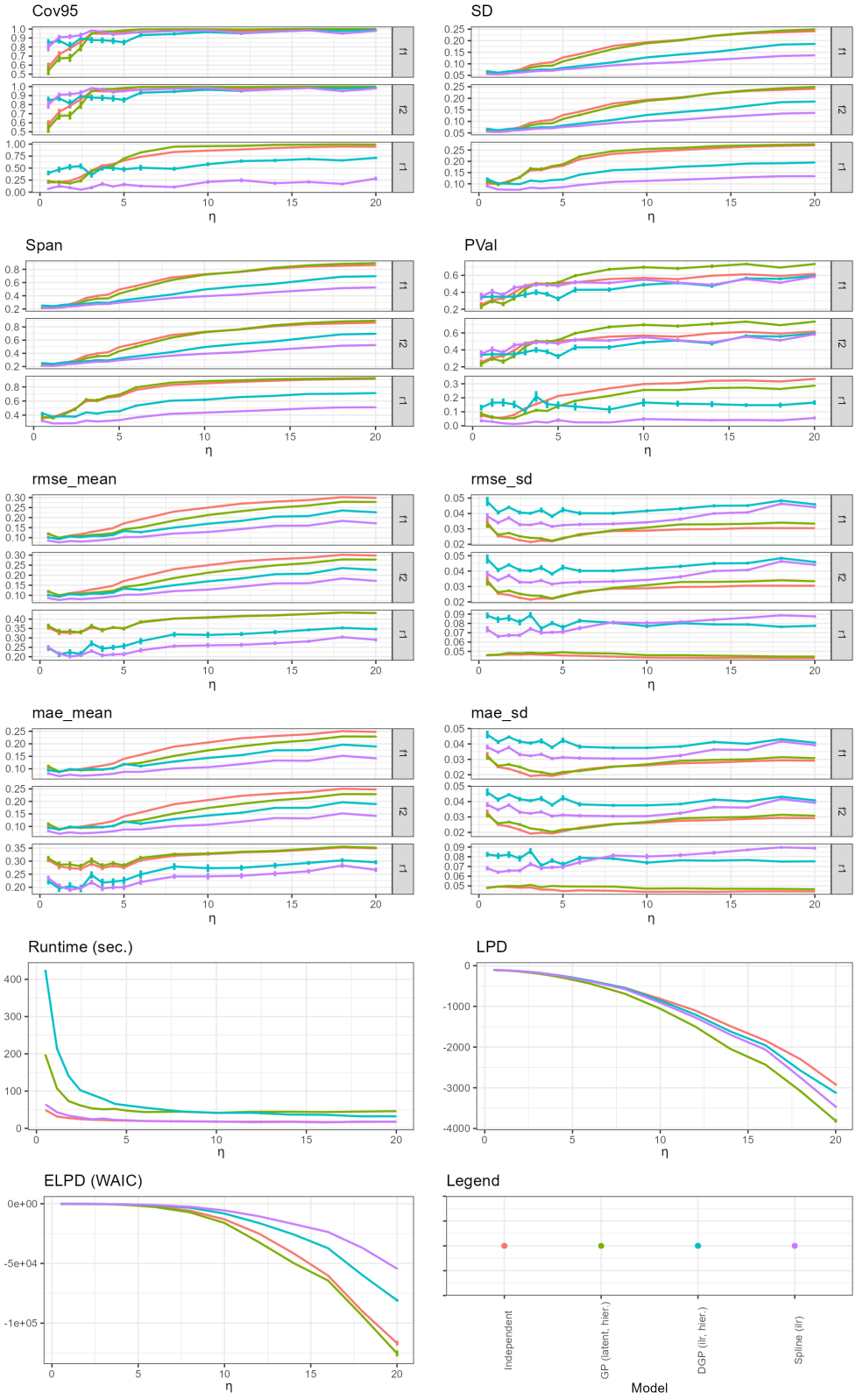

Figure B.6.: (b) Influence of Measurement Noise (subsection 4.3.4).

Appendix C

Application to Datasets

C.1. Monolith Dataset

The monolith dataset contains 16 time series of $\delta^{15}N$ and $\delta^{15}N^{SP}$ measurements split into five classes of monoliths. It was collected by Harris, Diaz-Pines, Stoll, *et al.* [4] for the purpose of measuring the dynamics of nitrification and denitrification after fertilization and in conjunction with drought and different precipitation conditions.

The measurements are plotted per class with colors indicating each separate time series in Figure C.1. Values tend to be large with correspondingly large uncertainties.

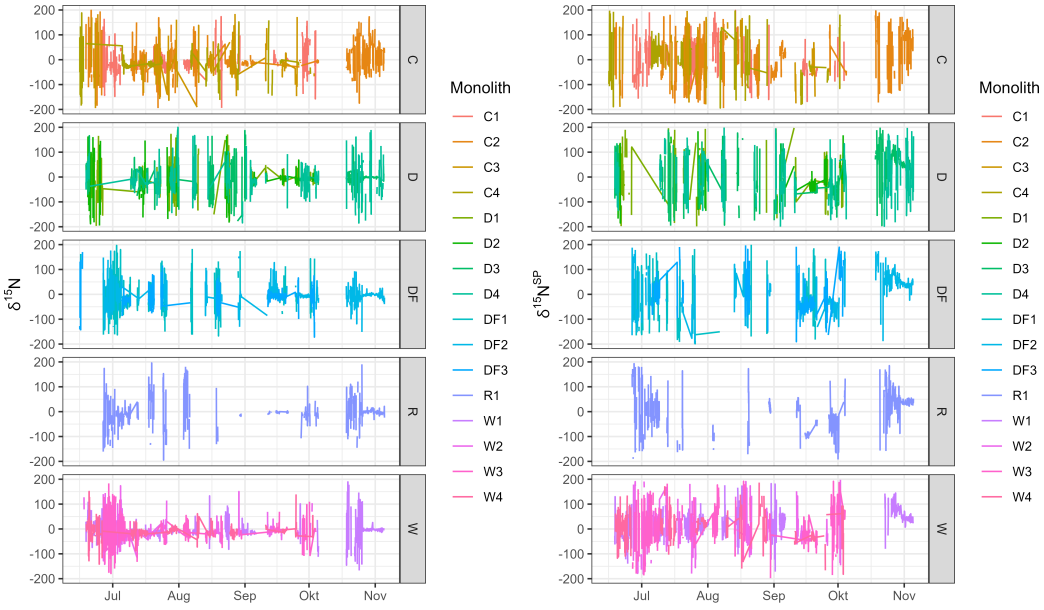

Figure C.1.: Monolith time series data of two isotopic measurements split into five categories of monoliths. The individual measurement series are marked with separate colors.

Dual isotope plots in Figure C.2 show the wide spread of the measurements. Nitrification and bacterial denitrification sources are shown as grey rectangles for reference. It seems like the measurements show movement over time starting off with very low $\delta^{15}N^{SP}$ values before concentrating into a more reasonable range towards the end of the experiment days.

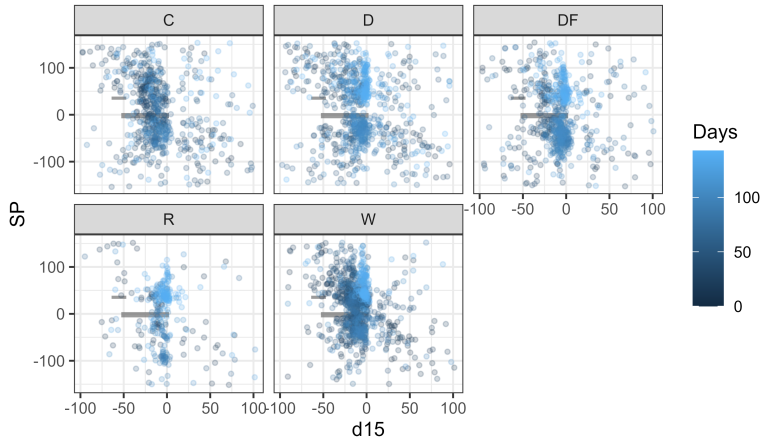

Figure C.2.: Dual isotope plot of the isotopic measurements for the five monolith classes with $\delta^{15}\mathrm{N}$ measurements on the x-axis and $\delta^{15}\mathrm{N}^{\mathrm{SP}}$ measurements on the y-axis. Different time points are separated with the marker color.

The plot of the logarithmic standard deviations involved in Figure C.3 shows that they differ by at least one order of magnitude at each given point in time. Especially towards the end the values are significantly lower.

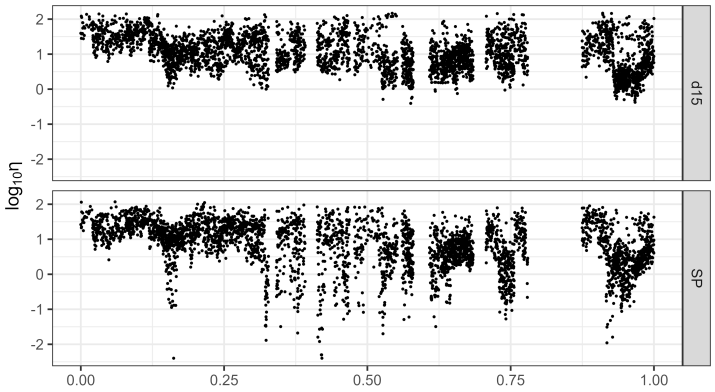

Figure C.3.: Log measurement noise over normalized experiment time aggregated over all time series simultaneously.

The raw measurements are very noisy but fortunately there are sufficiently many. This means that they can be grouped together by aggregation and smoothing methods. In addition to classical binning or moving average computations, kernel smoothing and spline smoothing could be applied to estimate less noisy mean time series. However, incorporating known standard deviations and estimating uncertainty of the smoothed time series is hard with all these methods. Local linear regression offers a way of quantifying the uncertainty of the smoothing estimator as well as weighting the raw measurements by inverse variance.

Shown are the raw measurements with their inverse variance weight indicated by size in Figure C.4. The blue dots are estimated smoothed values with standard errors shown as error bars. Additionally, the values are estimated only at every third raw measurement point. This reduces the number of points that have to be processed by following models and increases certainty of estimation by grouping points together.

Now the time series models can be run on the smoothed time series. The independent time step model, the generalized linear model with splines as well as the Dirichlet-Gaussian process prior is examined. Degrees of freedom are 8 for the source contributions and 4 for fractionation. Correlation lengths are set to 1% of the total experiment duration for the source contributions and 10% of the total experiment duration for fractionation.

Estimated source contributions and fractionation are noisy as well as can be seen in Figure C.5 and Figure C.6. The Dirichlet-Gaussian process prior on source contributions seems to be only marginally influential due to the short correlation length, but coupling fractionation improves posterior variances. The effective sample size N_{eff} gives an indication of how accurate the posterior means and variance can be expected to be and the Gelman-Rubin statistic $\hat{R}$ gives an indication of how close the sampled points are to a stationary sample.

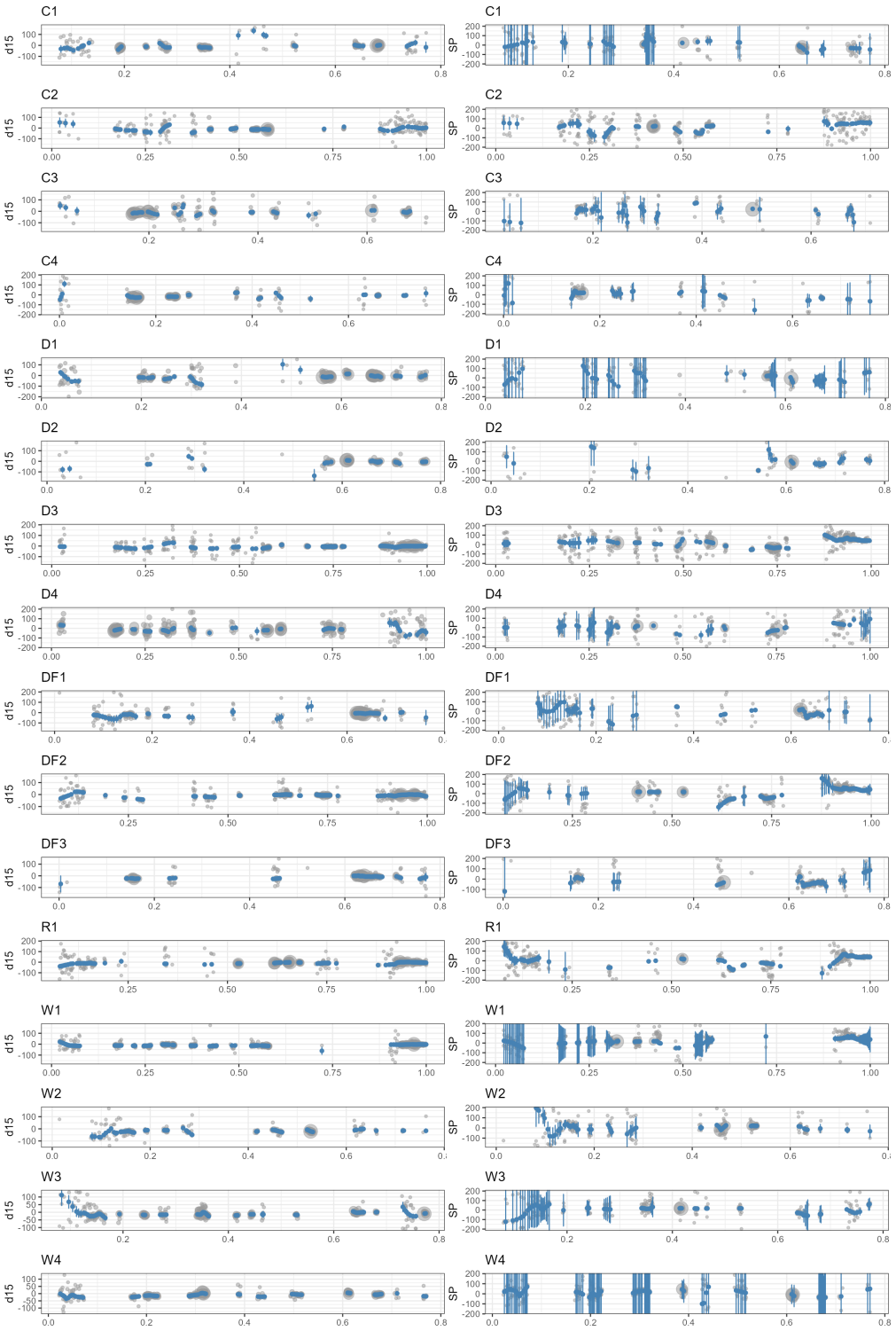

Figure C.4.: LOESS smoothed time series with uncertainty estimation shown as vertical bars. The original time series is shown as gray dots with size corresponding to inverse variance weights.

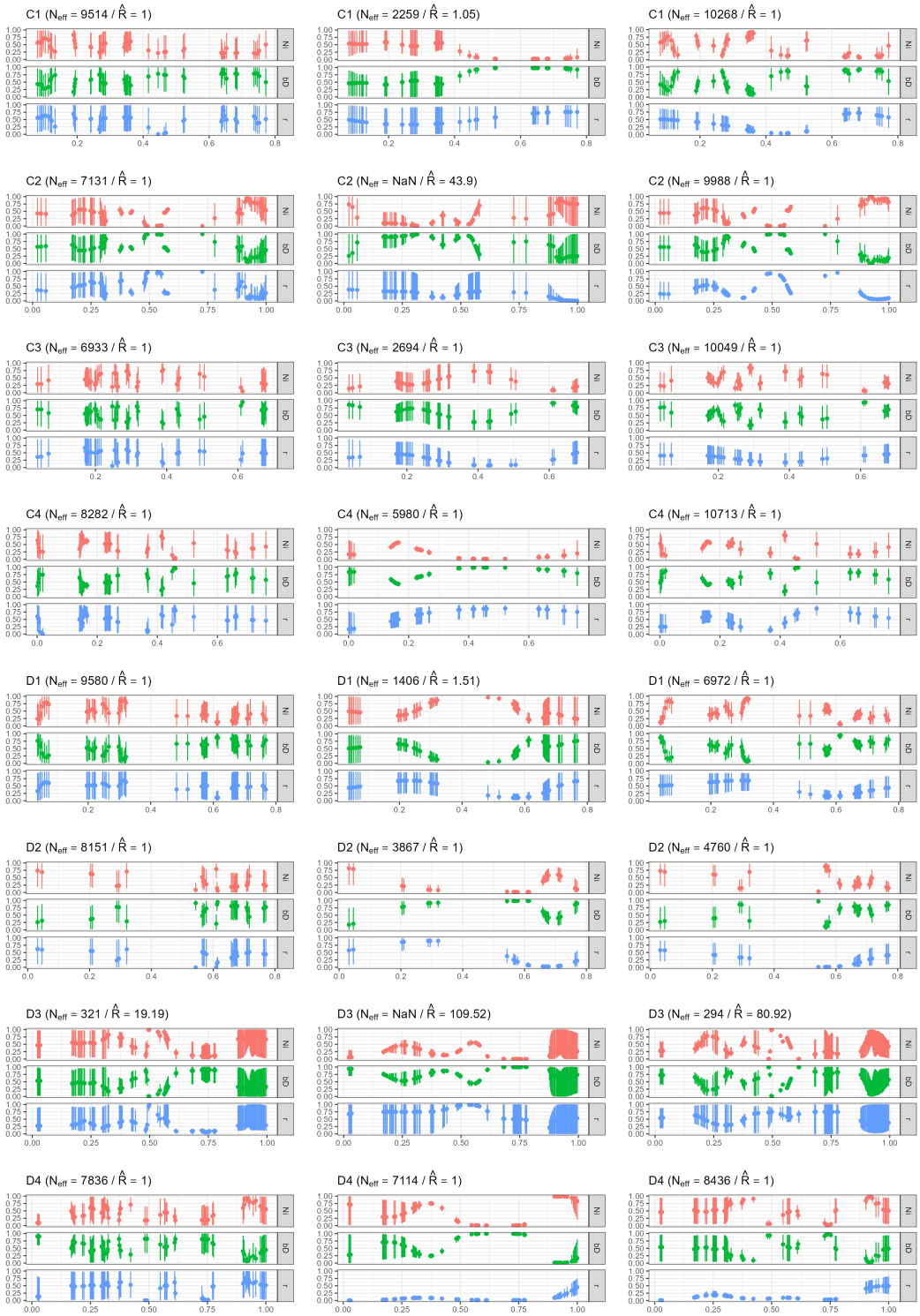

Figure C.5.: (a) Results of independent time step model (left), generalized linear model with spline basis (middle) and Dirichlet-Gaussian process prior (right).

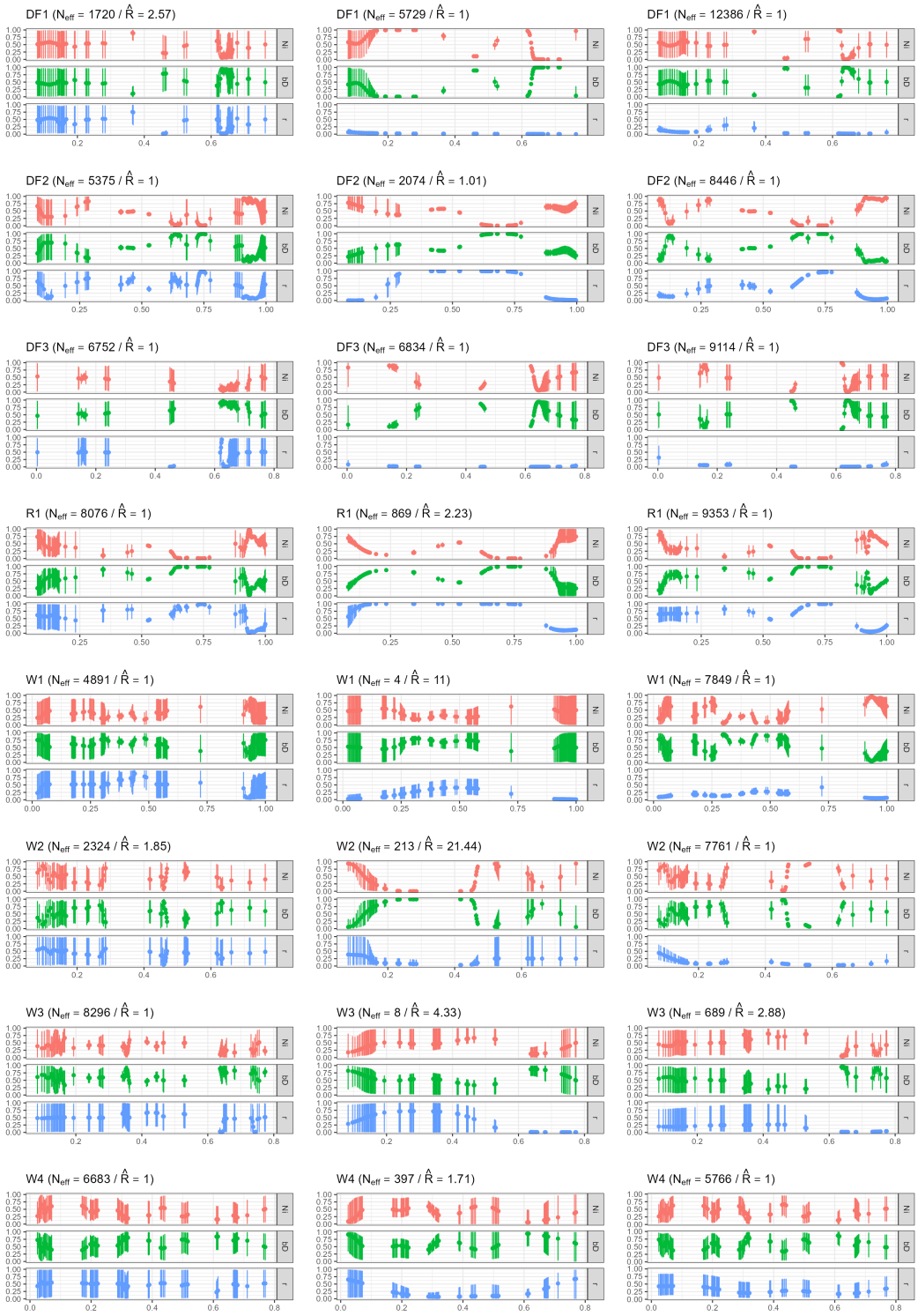

Figure C.6.: (b) Results of independent time step model (left), generalized linear model with spline basis (middle) and Dirichlet-Gaussian process prior (right).

C.2. Climate Manipulation Dataset

This dataset consists of 12 time series of three isotopic measurements δ^{15}N, $\delta^{15}N^{SP}$ and δ^{18}O collected by Stoll, Diaz-Pines, Reinthaler, *et al.* [112]. The goal of the experiment is to study managed grassland in a manipulated climate that simulates extreme droughts in terms of temperature, atmospheric gasses and precipitation. Measurement values span a wide range and are equipped with estimates for their standard deviation.

In order to get an estimate for the smoothed time series as well as quantify its uncertainty, local linear regression with inverse variance weights on the raw data points is used. The smoothing window is chosen to fit 10% of data points.

The source contributions are then estimated using the TimeFRAME models. The independent time step model, the generalized linear model with B-spline basis and the Dirichlet-Gaussian prior model are chosen for this case. Spline parameters are left to the default and correlation lengths are chosen to be 20% for source contributions and 5% for fractionation. The model fit is then examined for the posterior means of measurements shown in Figure C.9 and Figure C.10. This corresponds to a smoothing operation that is aware of the sources and the mixing equation. It seems like the δ^{18}O measurements are still out of proportion.

The estimates for source contributions and fractionation are plotted as mean lines with 95% confidence intervals in Figure C.11 and Figure C.12. Effective sample sizes N_{eff} give an indication of how reliable the posterior estimates are and the Gelman-Rubin statistic $\hat{R}$ indicates good chain mixing and convergence to the stationary distribution. Most areas of the time series seem to be uninformative, although some areas have very strong estimates.

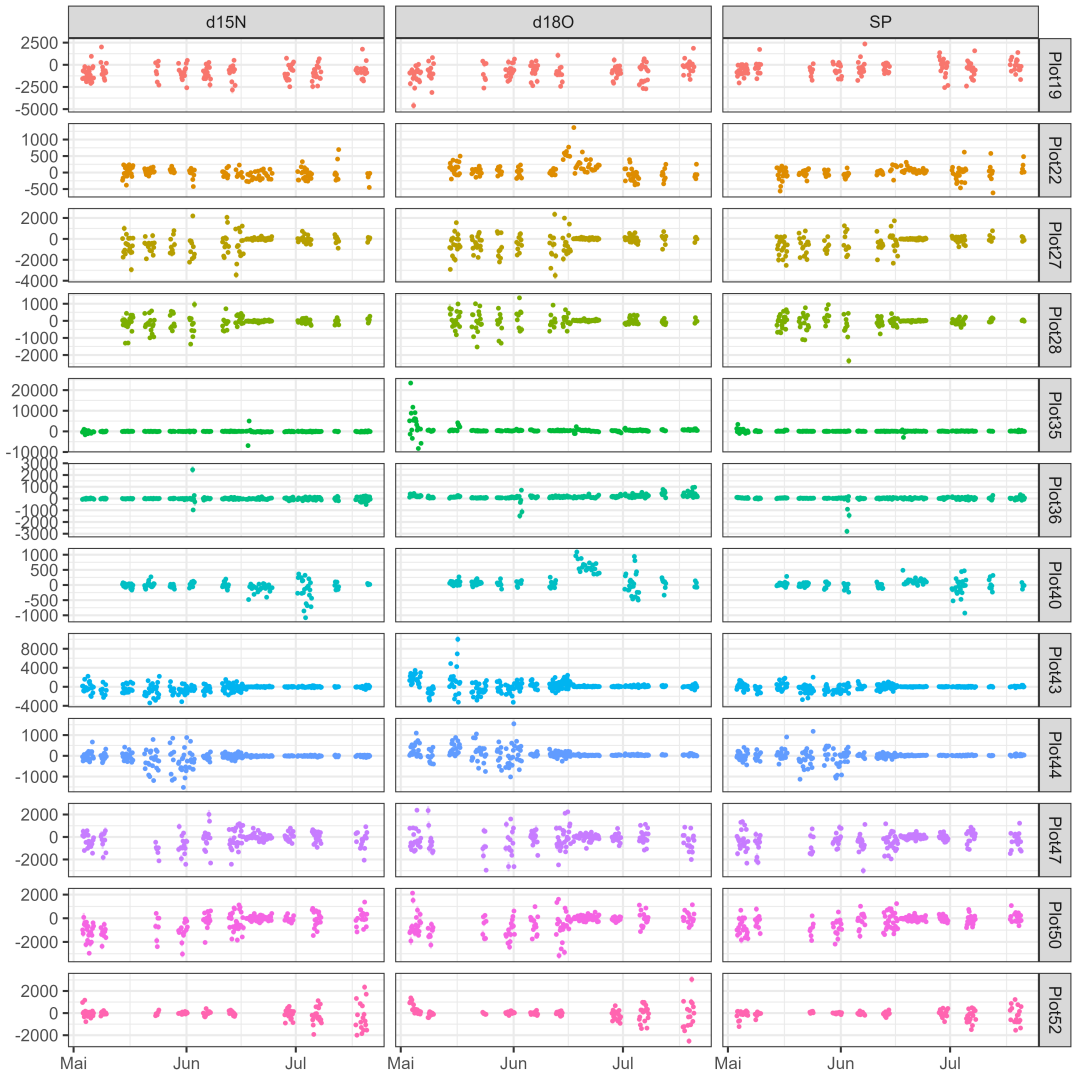

Figure C.7.: Isotopic measurements of all time series included in the dataset.

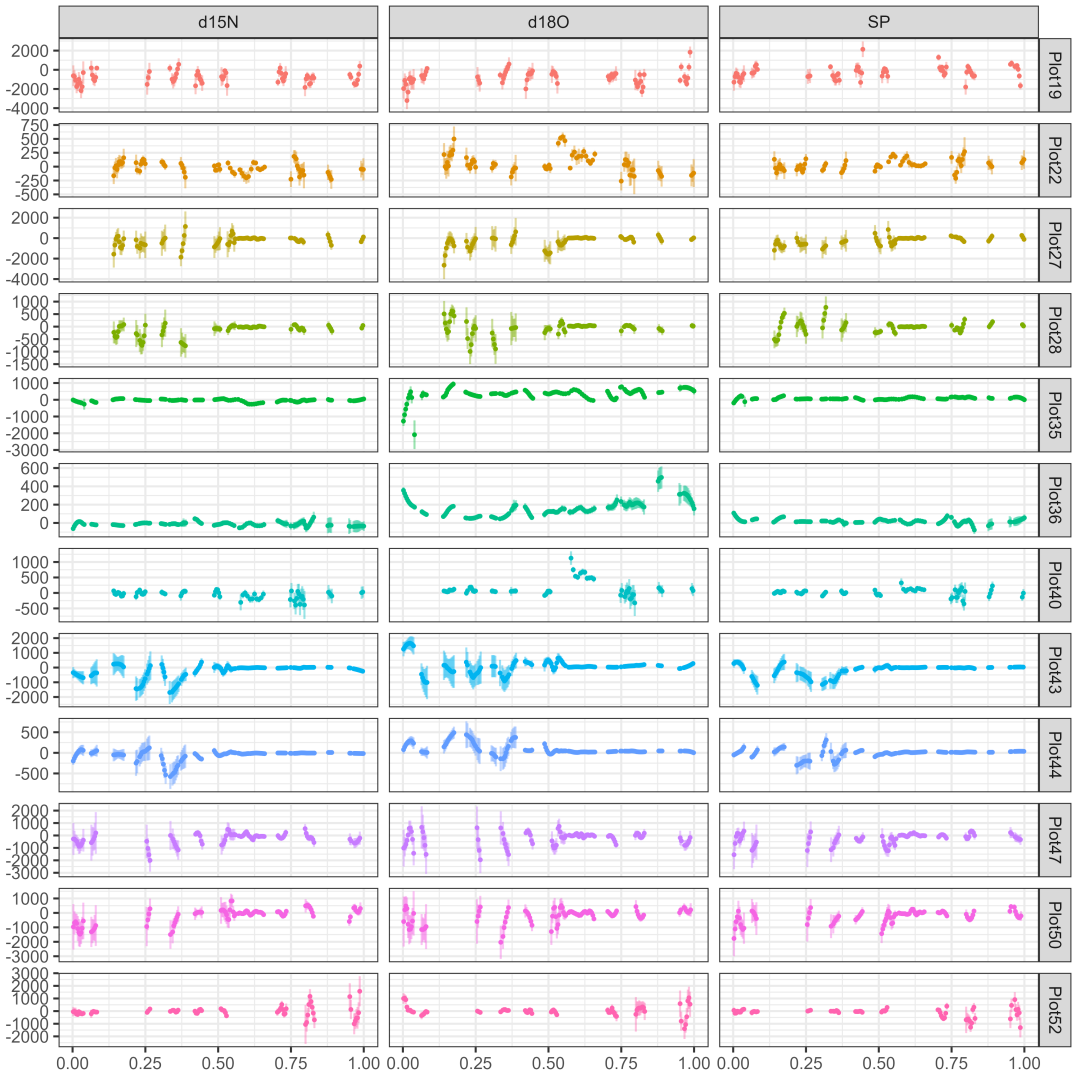

Figure C.8.: LOESS fits of all times series and isotopic measurements.

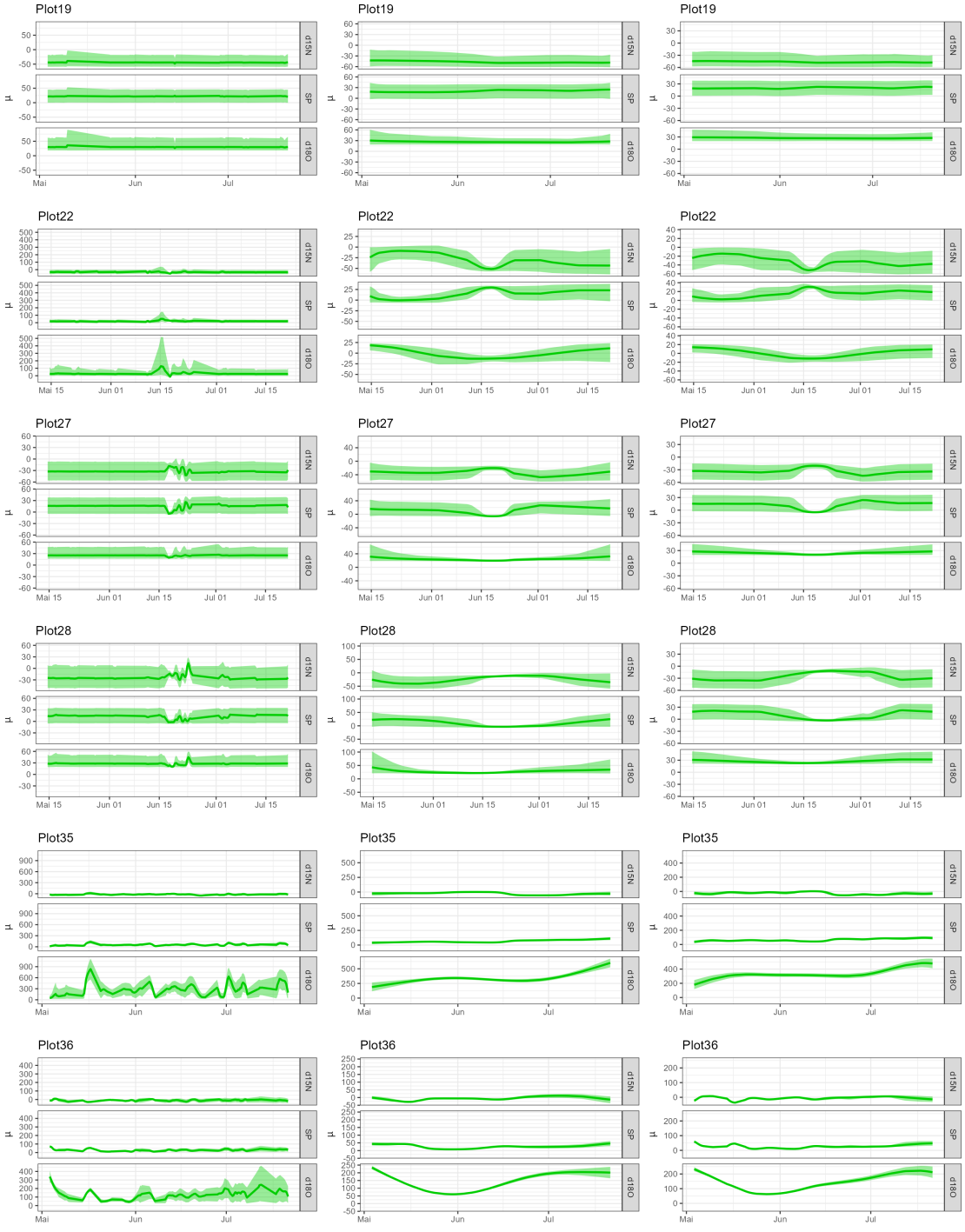

Figure C.9.: (a) Posterior estimates of the isotopic measurement mean for all 12 time series and all isotopic measurements for the independent time step model (left), spline-based model (middle) and Dirichlet-Gaussian process model (right).

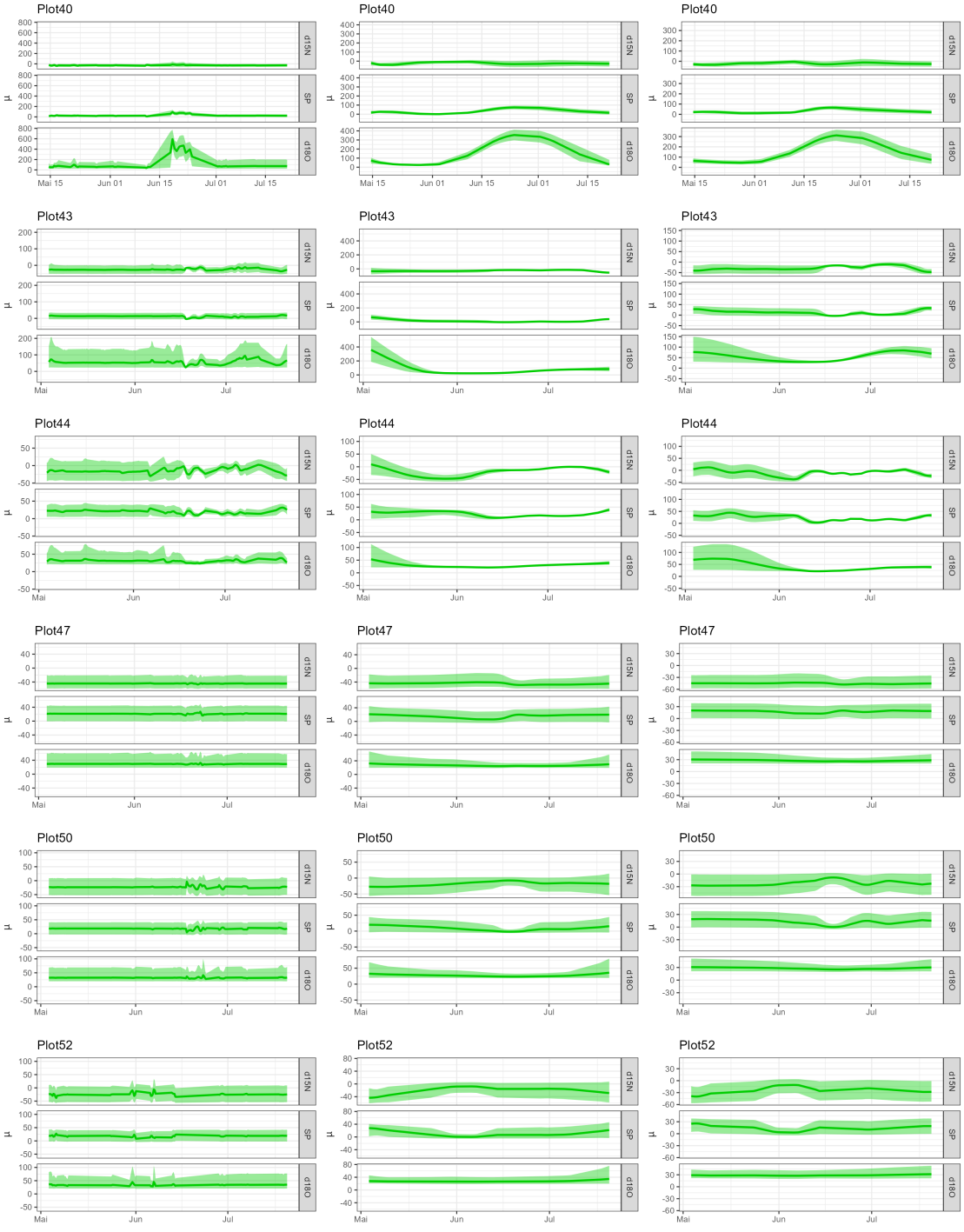

Figure C.10.: (b) Posterior estimates of the isotopic measurement mean for all 12 time series and all isotopic measurements for the independent time step model (left), spline-based model (middle) and Dirichlet-Gaussian process model (right).

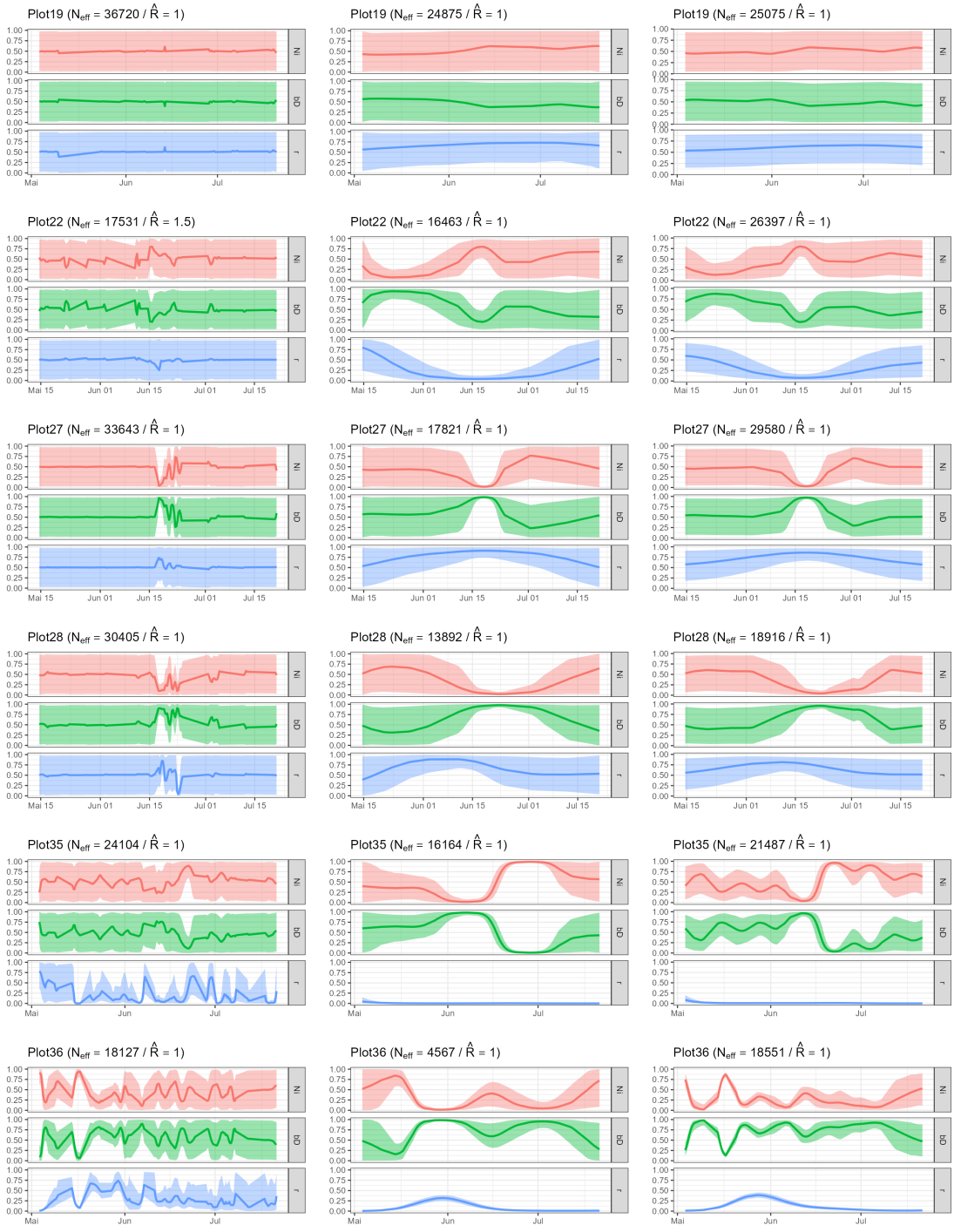

Figure C.11.: (a) Estimates of source contributions and fractionation for the independent time step model (left), spline-based model (middle) and Dirichlet-Gaussian process model (right).

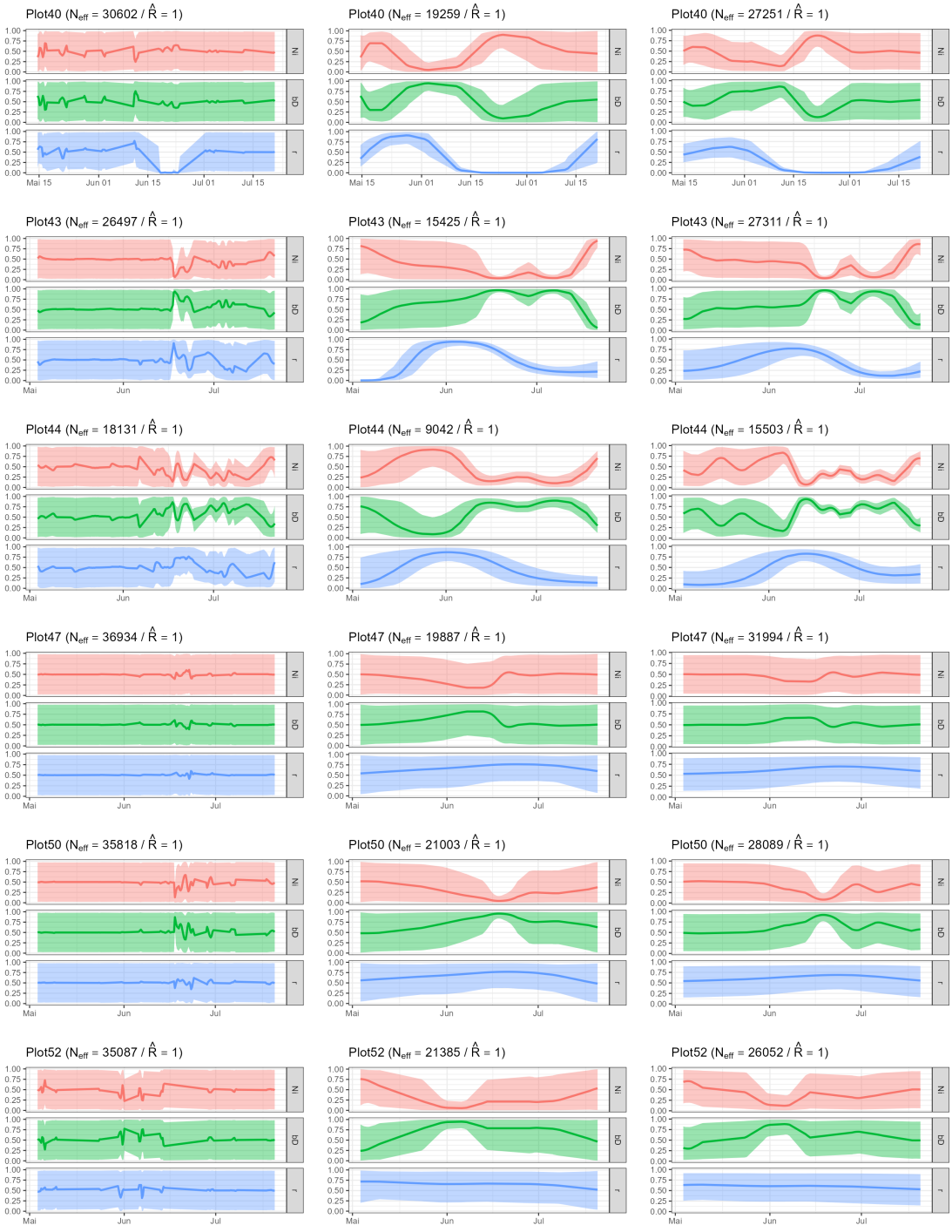

Figure C.12.: (b) Estimates of source contributions and fractionation for the independent time step model (left), spline-based model (middle) and Dirichlet-Gaussian process model (right).

Bibliography

[1] D. Xue, J. Botte, B. D. Baets, *et al.*, "Present limitations and future prospects of stable isotope methods for nitrate source identification in surface- and groundwater," *Water Research*, vol. 43, no. 5, pp. 1159–1170, Mar. 2009. DOI: 10.1016/j.watres.2008.12.048. [Online]. Available: https://doi.org/10.1016/j.watres.2008.12.048.

[2] D. L. Phillips, "Converting isotope values to diet composition: The use of mixing models," *Journal of Mammalogy*, vol. 93, no. 2, pp. 342–352, Apr. 2012. DOI: 10.1644/11-mamm-s-158.1. [Online]. Available: https://doi.org/10.1644/11-mamm-s-158.1.

[3] R. L. Sutka, N. E. Ostrom, P. H. Ostrom, *et al.*, "Distinguishing nitrous oxide production from nitrification and denitrification on the basis of isotopomer abundances," *Applied and Environmental Microbiology*, vol. 72, no. 1, pp. 638–644, Jan. 2006. DOI: 10.1128/aem.72.1.638-644.2006. [Online]. Available: https://doi.org/10.1128/aem.72.1.638-644.2006.

[4] E. Harris, E. Diaz-Pines, E. Stoll, *et al.*, "Denitrifying pathways dominate nitrous oxide emissions from managed grassland during drought and rewetting," *Science Advances*, vol. 7, no. 6, Feb. 2021. DOI: 10.1126/sciadv.abb7118. [Online]. Available: https://doi.org/10.1126/sciadv.abb7118.

[5] K. Butterbach-Bahl, E. M. Baggs, M. Dannenmann, R. Kiese, and S. Zechmeister-Boltenstern, "Nitrous oxide emissions from soils: How well do we understand the processes and their controls?" *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 368, no. 1621, p. 20 130 122, Jul. 2013. DOI: 10.1098/rstb.2013.0122. [Online]. Available: https://doi.org/10.1098/rstb.2013.0122.

[6] E. Harris, A. Joss, L. Emmenegger, *et al.*, "Isotopic evidence for nitrous oxide production pathways in a partial nitritation-anammox reactor," *Water Research*, vol. 83, pp. 258–270, Oct. 2015. DOI: 10.1016/j.watres.2015.06.040. [Online]. Available: https://doi.org/10.1016/j.watres.2015.06.040.

[7] H. Tian, J. Yang, R. Xu, *et al.*, "Global soil nitrous oxide emissions since the preindustrial era estimated by an ensemble of terrestrial biosphere models: Magnitude, attribution, and uncertainty," *Global Change Biology*, vol. 25, no. 2, pp. 640–659, Dec. 2018. DOI: 10.1111/gcb.14514. [Online]. Available: https://doi.org/10.1111/gcb.14514.

[8] K. Thoning, E. Dlugokencky, X. Lan, and NOAA Global Monitoring Laboratory, *Trends in globally-averaged CH4, N2O, and SF6*, 2022. DOI: 10.15138/P8XG-AA10. [Online]. Available: https://gml.noaa.gov/ccgg/trends_doi.html.

[9] S. A. Montzka, E. J. Dlugokencky, and J. H. Butler, "Non-CO2 greenhouse gases and climate change," *Nature*, vol. 476, no. 7358, pp. 43–50, Aug. 2011. DOI: 10.1038/nature10322. [Online]. Available: https://doi.org/10.1038/nature10322.

[10] A. R. Ravishankara, J. S. Daniel, and R. W. Portmann, "Nitrous oxide (N_2O): The dominant ozone-depleting substance emitted in the 21st century," *Science*, vol. 326, no. 5949, pp. 123–125, Oct. 2009. DOI: 10.1126/science.1176985. [Online]. Available: https://doi.org/10.1126/science.1176985.

[11] U. Purkhold, A. Pommerening-Röser, S. Juretschko, M. C. Schmid, H.-P. Koops, and M. Wagner, "Phylogeny of all recognized species of ammonia oxidizers based on comparative 16s rRNA and iamoA/i sequence analysis: Implications for molecular diversity surveys," *Applied and Environmental Microbiology*, vol. 66, no. 12, pp. 5368–5382, Dec. 2000. DOI: 10.1128/aem.66.12.5368-5382.2000. [Online]. Available: https://doi.org/10.1128/aem.66.12.5368-5382.2000.

[12] W. G. Zumft, "Cell biology and molecular basis of denitrification," *Microbiology and Molecular Biology Reviews*, vol. 61, no. 4, pp. 533–616, Dec. 1997. DOI: 10.1128/mmbr.61.4.533-616.1997. [Online]. Available: https://doi.org/10.1128/mmbr.61.4.533-616.1997.

[13] W. M. Haynes, D. R. Lide, and T. J. Bruno, *CRC Handbook of Chemistry and Physics (95th ed.)* CRC Press, 2014.

[14] S. Toyoda, M. Yano, S.-i. Nishimura, *et al.*, "Characterization and production and consumption processes of nsub2/subo emitted from temperate agricultural soils determined via isotopomer ratio analysis," *Global Biogeochemical Cycles*, vol. 25, no. 2, May 2011. DOI: 10.1029/2009gb003769. [Online]. Available: https://doi.org/10.1029/2009gb003769.

[15] D. Wu, R. Well, L. M. Cárdenas, *et al.*, "Quantifying N_2O reduction to N_2 during denitrification in soils via isotopic mapping approach: Model evaluation and uncertainty analysis," *Environmental Research*, vol. 179, p. 108 806, Dec. 2019. DOI: 10.1016/j.envres.2019.108806. [Online]. Available: https://doi.org/10.1016/j.envres.2019.108806.

[16] D. L. Phillips and J. W. Gregg, "Source partitioning using stable isotopes: Coping with too many sources," *Oecologia*, vol. 136, no. 2, pp. 261–269, Jul. 2003. DOI: 10.1007/s00442-003-1218-3. [Online]. Available: https://doi.org/10.1007/s00442-003-1218-3.

[17] D. Lewicka-Szczebak, M. P. Lewicki, and R. Well, "N_2O isotope approaches for source partitioning of N_2O production and estimation of N_2O reduction – validation with the ^{15}n gas-flux method in laboratory and field studies," *Biogeosciences*, vol. 17, no. 22, pp. 5513–5537, Nov. 2020. DOI: 10.5194/bg-17-5513-2020. [Online]. Available: https://doi.org/10.5194/bg-17-5513-2020.

[18] M. P. Lewicki, D. Lewicka-Szczebak, and G. Skrzypek, "FRAME—monte carlo model for evaluation of the stable isotope mixing and fractionation," *PLOS ONE*, vol. 17, no. 11, V. Kovtun, Ed., e0277204, Nov. 2022. DOI: 10.1371/journal.pone.0277204. [Online]. Available: https://doi.org/10.1371/journal.pone.0277204.

[19] M. P. Lewicki, FRAME*: Fractionation and mixing evaluation of isotopes*, Accessed: 2023-05-28, 2021. [Online]. Available: https://github.com/malewick/frame/.

[20] B. C. Stock, A. L. Jackson, E. J. Ward, A. C. Parnell, D. L. Phillips, and B. X. Semmens, "Analyzing mixing systems using a new generation of bayesian tracer mixing models," *PeerJ*, vol. 6, e5096, Jun. 2018. DOI: 10.7717/peerj.5096. [Online]. Available: https://doi.org/10.7717/peerj.5096.

[21] A. C. Parnell, D. L. Phillips, S. Bearhop, *et al.*, "Bayesian stable isotope mixing models," *Environmetrics*, pp. 387–399, Jul. 2013. DOI: 10.1002/env.2221. [Online]. Available: https://doi.org/10.1002/env.2221.

[22] B. X. Semmens, E. J. Ward, J. W. Moore, and C. T. Darimont, "Quantifying inter- and intra-population niche variability using hierarchical bayesian stable isotope mixing models," *PLoS ONE*, vol. 4, no. 7, W. M. Getz, Ed., e6187, Jul. 2009. DOI: 10.1371/journal.pone.0006187. [Online]. Available: https://doi.org/10.1371/journal.pone.0006187.

[23] J. Hayes, "An introduction to isotopic calculations," Tech. Rep., Sep. 2004. DOI: 10.1575/1912/27058. [Online]. Available: https://doi.org/10.1575/1912/27058.

[24] C. R. McKinney, J. M. McCrea, S. Epstein, H. A. Allen, and H. C. Urey, "Improvements in mass spectrometers for the measurement of small differences in isotope abundance ratios," *Review of Scientific Instruments*, vol. 21, no. 8, pp. 724–730, Aug. 1950. DOI: 10.1063/1.1745698. [Online]. Available: https://doi.org/10.1063/1.1745698.

[25] F. Maggi and W. J. Riley, "Mathematical treatment of isotopologue and isotopomer speciation and fractionation in biochemical kinetics," *Geochimica et Cosmochimica Acta*, vol. 74, no. 6, pp. 1823–1835, Mar. 2010. DOI: 10.1016/j.gca.2009.12.021. [Online]. Available: https://doi.org/10.1016/j.gca.2009.12.021.

[26] T. R. Denk, J. Mohn, C. Decock, *et al.*, "The nitrogen cycle: A review of isotope effects and isotope modeling approaches," *Soil Biology and Biochemistry*, vol. 105, pp. 121–137, Feb. 2017. DOI: 10.1016/j.soilbio.2016.11.015. [Online]. Available: https://doi.org/10.1016/j.soilbio.2016.11.015.

[27] R. Penrose, "On best approximate solutions of linear matrix equations," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 52, no. 1, pp. 17–19, Jan. 1956. DOI: 10.1017/s0305004100030929. [Online]. Available: https://doi.org/10.1017/s0305004100030929.

[28] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses*. Springer-Verlag, 2003. DOI: 10.1007/b97366. [Online]. Available: https://doi.org/10.1007/b97366.

[29] D. Lemons and P. Langevin, *An Introduction to Stochastic Processes in Physics* (Johns Hopkins Paperback). Johns Hopkins University Press, 2002, ISBN: 9780801868672.

[30] A. O'Hagan and J. J. Forster, *Kendall's Advanced Theory of Statistics, volume 2B: Bayesian Inference, second edition*. New York: Oxford University Press, 2004, vol. 2B.

[31] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis, Third Edition* (Chapman & Hall/CRC Texts in Statistical Science). Taylor & Francis, 2013, ISBN: 9781439840955.

[32] A. Gelman, A. Vehtari, D. Simpson, *et al.*, "Bayesian workflow," 2020. DOI: 10.48550/ARXIV.2011.01808. [Online]. Available: https://arxiv.org/abs/2011.01808.

[33] J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman, "Visualization in bayesian workflow," *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 182, no. 2, pp. 389–402, Jan. 2019. DOI: 10.1111/rssa.12378. [Online]. Available: https://doi.org/10.1111/rssa.12378.

[34] P. H. Garthwaite, J. B. Kadane, and A. O'Hagan, "Statistical methods for eliciting probability distributions," *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 680–701, Jun. 2005. DOI: 10.1198/016214505000000105. [Online]. Available: https://doi.org/10.1198/016214505000000105.

[35] R. Yang and J. O. Berger, "A catalog of noninformative priors," 1996.

[36] B. P. Carlin and T. A. Louis, "Empirical bayes: Past, present and future," *Journal of the American Statistical Association*, vol. 95, no. 452, pp. 1286–1289, 2000. DOI: 10.1080/01621459.2000.10474331.

[37] C. N. Morris, "Parametric empirical bayes inference: Theory and applications," *Journal of the American Statistical Association*, vol. 78, no. 381, pp. 47–55, Mar. 1983. DOI: 10.1080/01621459.1983.10477920. [Online]. Available: https://doi.org/10.1080/01621459.1983.10477920.

[38] G. M. Allenby, P. E. Rossi, and R. E. McCulloch, "Hierarchical bayes models: A practitioners guide," *SSRN Electronic Journal*, 2005. DOI: 10.2139/ssrn.655541. [Online]. Available: https://doi.org/10.2139/ssrn.655541.

[39] N. P. Lemoine, "Moving beyond noninformative priors: Why and how to choose weakly informative priors in bayesian analyses," *Oikos*, vol. 128, no. 7, pp. 912–928, Apr. 2019. DOI: 10.1111/oik.05985. [Online]. Available: https://doi.org/10.1111/oik.05985.

[40] M. S. deVries, B. C. Stock, J. H. Christy, G. R. Goldsmith, and T. E. Dawson, "Specialized morphology corresponds to a generalist diet: Linking form and function in smashing mantis shrimp crustaceans," *Oecologia*, vol. 182, no. 2, pp. 429–442, Jun. 2016. DOI: 10.1007/s00442-016-3667-5. [Online]. Available: https://doi.org/10.1007/s00442-016-3667-5.

[41] A. Gelman, D. Simpson, and M. Betancourt, "The prior can often only be understood in the context of the likelihood," *Entropy*, vol. 19, no. 10, p. 555, Oct. 2017. DOI: 10.3390/e19100555. [Online]. Available: https://doi.org/10.3390/e19100555.

[42] C. Röver, R. Bender, S. Dias, *et al.*, "On weakly informative prior distributions for the heterogeneity parameter in bayesian random-effects meta-analysis," *Research Synthesis Methods*, Feb. 2021. DOI: 10.1002/jrsm.1475. [Online]. Available: https://doi.org/10.1002/jrsm.1475.

[43] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 186, no. 1007, pp. 453–461, Sep. 1946. DOI: 10.1098/rspa.1946.0056. [Online]. Available: https://doi.org/10.1098/rspa.1946.0056.

[44] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. Springer-Verlag, 1998. DOI: 10.1007/b98854. [Online]. Available: https://doi.org/10.1007/b98854.

[45] C. P. Robert, N. Chopin, and J. Rousseau, "Harold Jeffreys's theory of probability revisited," *Statistical Science*, vol. 24, no. 2, May 2009. DOI: 10.1214/09-sts284. [Online]. Available: https://doi.org/10.1214/09-sts284.

[46] E. Jaynes, "Prior probabilities," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 3, pp. 227–241, 1968. DOI: 10.1109/tssc.1968.300117. [Online]. Available: https://doi.org/10.1109/tssc.1968.300117.

[47] N. C. Metropolis, "The beginning of the monte carlo method," *Los Alamos Science*, pp. 125–130, 1987.

[48] G. E. Forsythe, "Von neumann's comparison method for random sampling from the normal and other distributions," *Mathematics of Computation*, vol. 26, no. 120, p. 817, Oct. 1972. DOI: 10.2307/2005864. [Online]. Available: https://doi.org/10.2307/2005864.

[49] G. Casella, C. P. Robert, and M. T. Wells, "Generalized accept-reject sampling schemes," in *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, Institute of Mathematical Statistics, 2004, pp. 342–347. DOI: 10.1214/lnms/1196285403. [Online]. Available: https://doi.org/10.1214/lnms/1196285403.

[50] T. Kloek and H. K. van Dijk, "Bayesian estimates of equation system parameters: An application of integration by monte carlo," *Econometrica*, vol. 46, no. 1, p. 1, Jan. 1978. DOI: 10.2307/1913641. [Online]. Available: https://doi.org/10.2307/1913641.

[51] D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo*. Chapman and Hall/CRC, May 2006. DOI: 10.1201/9781482296426. [Online]. Available: https://doi.org/10.1201/9781482296426.

[52] C. Sherlock, P. Fearnhead, and G. O. Roberts, "The random walk metropolis: Linking theory and practice through a case study," 2010. DOI: 10.48550/ARXIV.1011.6217. [Online]. Available: https://arxiv.org/abs/1011.6217.

[53] C. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.

[54] S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid monte carlo," *Physics Letters B*, vol. 195, no. 2, pp. 216–222, Sep. 1987. DOI: 10.1016/0370-2693(87)91197-x. [Online]. Available: https://doi.org/10.1016/0370-2693(87)91197-x.

[55] M. D. Homan and A. Gelman, "The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1593–1623, Jan. 2014, ISSN: 1532-4435.

[56] M. J. Betancourt, "Generalizing the no-u-turn sampler to riemannian manifolds," 2013. DOI: 10.48550/ARXIV.1304.1920. [Online]. Available: https://arxiv.org/abs/1304.1920.

[57] D. McQuarrie, *Statistical Mechanics* (G - Reference,Information and Interdisciplinary Subjects Series). University Science Books, 2000, ISBN: 9781891389153.

[58] P. Diaconis and L. Saloff-Coste, "What do we know about the metropolis algorithm?" *Journal of Computer and System Sciences*, vol. 57, no. 1, pp. 20–36, Aug. 1998. DOI: 10.1006/jcss.1998.1576. [Online]. Available: https://doi.org/10.1006/jcss.1998.1576.

[59] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers, 7th Edition*. Wiley, 2018, ISBN: 9781119456261.

[60] C. J. Geyer, *Introduction to Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011, ISBN: 9780429138508.

[61] G. L. Jones, "On the markov chain central limit theorem," *Probability Surveys*, vol. 1, Jan. 2004. DOI: 10.1214/154957804100000051. [Online]. Available: https://doi.org/10.1214/154957804100000051.

[62] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, no. 4, Nov. 1992. DOI: 10.1214/ss/1177011136. [Online]. Available: https://doi.org/10.1214/ss/1177011136.

[63] S. P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations," *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, pp. 434–455, 1998. DOI: 10.1080/10618600.1998.10474787.

[64] D. Vats and C. Knudson, "Revisiting the Gelman-Rubin diagnostic," 2018. DOI: 10.48550/ARXIV.1812.09384.

[65] A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner, "Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC," 2019. DOI: 10.48550/ARXIV.1903.08008.

[66] Stan Development Team, *Stan modeling language users guide and reference manual*, Version 2.32, 2023. [Online]. Available: https://mc-stan.org/.

[67] M. Plummer, "JAGS: A program for analysis of bayesian graphical models using gibbs sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, 2003. [Online]. Available: https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf.

[68] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/.

[69] Stan Development Team, *RStan: The R interface to Stan*, R package version 2.21.8, 2023. [Online]. Available: https://mc-stan.org/.

[70] M. Plummer, A. Stukalov, and M. Denwood, *Rjags: Bayesian graphical models using MCMC*, R package version 4-14, 2023. [Online]. Available: https://cran.r-project.org/web/packages/rjags/index.html.

[71] E. A. Nadaraya, "On estimating regression," *Theory of Probability &amp Its Applications*, vol. 9, no. 1, pp. 141–142, Jan. 1964. DOI: 10.1137/1109020. [Online]. Available: https://doi.org/10.1137/1109020.

[72] I. Gijbels and I. Prosdocimi, "Loess," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 5, pp. 590–599, Jul. 2010. DOI: 10.1002/wics.104. [Online]. Available: https://doi.org/10.1002/wics.104.

[73] L. Schumaker, *Spline functions: basic theory*. Cambridge university press, 2007.

[74] A. Rencher and W. Christensen, *Methods of Multivariate Analysis* (Wiley Series in Probability and Statistics). Wiley, 2012, ISBN: 9781118391679.

[75] J. Rice and M. Rosenblatt, "Smoothing splines: Regression, derivatives and deconvolution," *The Annals of Statistics*, vol. 11, no. 1, pp. 141–156, 1983, ISSN: 00905364. [Online]. Available: http://www.jstor.org/stable/2240468.

[76] J.-P. Vert, K. Tsuda, and B. Schölkopf, "A primer on kernel methods," *Kernel methods in computational biology*, vol. 47, pp. 35–70, 2004.

[77] M. G. Genton, "Classes of kernels for machine learning: A statistics perspective," *J. Mach. Learn. Res.*, vol. 2, pp. 299–312, Mar. 2002, ISSN: 1532-4435.

[78] C. C. Lalescu, "Two hierarchies of spline interpolations. practical algorithms for multivariate higher order splines," 2009. DOI: 10.48550/ARXIV.0905.3564. [Online]. Available: https://arxiv.org/abs/0905.3564.

[79] A. Fick, "On liquid diffusion," *Journal of Membrane Science*, vol. 100, no. 1, pp. 33–38, Mar. 1995. DOI: 10.1016/0376-7388(94)00230-v. [Online]. Available: https://doi.org/10.1016/0376-7388(94)00230-v.

[80] K. Cole, J. Beck, A. Haji-Sheikh, and B. Litkouhi, *Heat Conduction Using Greens Functions*. CRC Press, Jul. 2010. DOI: 10.1201/9781439895214. [Online]. Available: https://doi.org/10.1201/9781439895214.

[81] C. Decock, J. Lee, M. Barthel, E. Verhoeven, F. Conen, and J. Six, "Process rate estimator: A novel model to predict total denitrification using natural abundance stable isotopes of nsub2/subo," Nov. 2022. DOI: 10.5194/bg-2022-221. [Online]. Available: https://doi.org/10.5194/bg-2022-221.

[82] J. Aitchison, "The statistical analysis of compositional data," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 44, no. 2, pp. 139–177, 1982, ISSN: 00359246. [Online]. Available: http://www.jstor.org/stable/2345821.

[83] V. Pawlowsky-Glahn, J. J. Egozcue, and R. Tolosana-Delgado, *Modelling and Analysis of Compositional Data*. John Wiley & Sons, Ltd, Mar. 2015. DOI: 10.1002/9781119003144. [Online]. Available: https://doi.org/10.1002/9781119003144.

[84] J. J. Egozcue, *Mathematical Geology*, vol. 35, no. 3, pp. 279–300, 2003. DOI: 10.1023/a:1023818214614. [Online]. Available: https://doi.org/10.1023/a:1023818214614.

[85] V. Nesrstová, I. Wilms, J. Palarea-Albaladejo, *et al.*, "Principal balances of compositional data for regression and classification using partial least squares," 2022. DOI: 10.48550/ARXIV.2211.01686. [Online]. Available: https://arxiv.org/abs/2211.01686.

[86] M. Greenacre, E. Grunsky, and J. Bacon-Shone, "A comparison of isometric and amalgamation logratio balances in compositional data analysis," *Computers &amp Geosciences*, vol. 148, p. 104 621, Mar. 2021. DOI: 10.1016/j.cageo.2020.104621. [Online]. Available: https://doi.org/10.1016/j.cageo.2020.104621.

[87] A. Parnell and R. Inger, *Stable isotope mixing models in r with simmr*, https://cran.r-project.org/web/packages/simmr/vignettes/simmr.html, Accessed: 2023-05-28.

[88] B. Stock and B. Semmens, *Mixsiar*, https://brianstock.github.io/MixSIAR/, Accessed: 2023-05-28.

[89] J. O. Berger, J. M. Bernardo, and D. Sun, "Overall objective priors," *Bayesian Analysis*, vol. 10, no. 1, Mar. 2015. DOI: 10.1214/14-ba915. [Online]. Available: https://doi.org/10.1214/14-ba915.

[90] K. B. Petersen and M. S. Pedersen, *The matrix cookbook*, Version 20121115, Nov. 2012. [Online]. Available: http://www2.compute.dtu.dk/pubdb/pubs/3274-full.html.

[91] A. B. Chan and D. Dong, "Generalized gaussian process models," in *CVPR 2011*, IEEE, Jun. 2011. DOI: 10.1109/cvpr.2011.5995688. [Online]. Available: https://doi.org/10.1109/cvpr.2011.5995688.

[92] J. Aitchison and S. M. Shen, "Logistic-normal distributions: Some properties and uses," *Biometrika*, vol. 67, no. 2, p. 261, Aug. 1980. DOI: 10.2307/2335470. [Online]. Available: https://doi.org/10.2307/2335470.

[93] L. Devroye, *Methods of Multivariate Analysis* (Wiley Series in Probability and Statistics). Wiley, 1986, ISBN: 9781118391679.

[94] A. B. Chan, "Multivariate generalized gaussian process models," 2013. DOI: 10.48550/ARXIV.1311.0360. [Online]. Available: https://arxiv.org/abs/1311.0360.

[95] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society. Series A (General)*, vol. 135, no. 3, p. 370, 1972. DOI: 10.2307/2344614. [Online]. Available: https://doi.org/10.2307/2344614.

[96] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. Springer US, 2021. DOI: 10.1007/978-1-0716-1418-1. [Online]. Available: https://doi.org/10.1007/978-1-0716-1418-1.

[97] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, Feb. 1970. DOI: 10.1080/00401706.1970.10488634. [Online]. Available: https://doi.org/10.1080/00401706.1970.10488634.

[98] W. J. Hemmerle, "An explicit solution for generalized ridge regression," *Technometrics*, vol. 17, no. 3, pp. 309–314, Aug. 1975. DOI: 10.1080/00401706.1975.10489333. [Online]. Available: https://doi.org/10.1080/00401706.1975.10489333.

[99] L. Yu, E. Harris, D. Lewicka-Szczebak, *et al.*, "What can we learn from N_2O isotope data? – analytics, processes and modelling," *Rapid Communications in Mass Spectrometry*, vol. 34, no. 20, Aug. 2020. DOI: 10.1002/rcm.8858. [Online]. Available: https://doi.org/10.1002/rcm.8858.

[100] W. Edwards, H. Lindman, and L. J. Savage, "Bayesian statistical inference for psychological research," in *Springer Series in Statistics*, Springer New York, 1992, pp. 531–578. DOI: 10.1007/978-1-4612-0919-5_34. [Online]. Available: https://doi.org/10.1007/978-1-4612-0919-5_34.

[101] D. B. Rubin, "Bayesianly justifiable and relevant frequency calculations for the applied statistician," *The Annals of Statistics*, vol. 12, no. 4, pp. 1151–1172, 1984, ISSN: 00905364, 21688966.

[102] A. Gelman, J. Hwang, and A. Vehtari, "Understanding predictive information criteria for bayesian models," 2013. DOI: 10.48550/ARXIV.1307.5928. [Online]. Available: https://arxiv.org/abs/1307.5928.

[103] A. Vehtari, A. Gelman, and J. Gabry, "Practical bayesian model evaluation using leave-one-out cross-validation and WAIC," *Statistics and Computing*, vol. 27, no. 5, pp. 1413–1432, Aug. 2016. DOI: 10.1007/s11222-016-9696-4. [Online]. Available: https://doi.org/10.1007/s11222-016-9696-4.

[104] J. Kruschke, *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Elsevier, 2014, ISBN: 9780124058880.

[105] M. Beraha, D. Falco, and A. Guglielmi, "JAGS, NIMBLE, Stan: A detailed comparison among bayesian MCMC software," 2021. DOI: 10.48550/ARXIV.2107.09357. [Online]. Available: https://arxiv.org/abs/2107.09357.

[106] R. J. Hyndman and H. L. Shang, "Rainbow plots, bagplots, and boxplots for functional data," *Journal of Computational and Graphical Statistics*, vol. 19, no. 1, pp. 29–45, Jan. 2010. DOI: 10.1198/jcgs.2009.08158. [Online]. Available: https://doi.org/10.1198/jcgs.2009.08158.

[107] Y. Sun and M. G. Genton, "Functional boxplots," *Journal of Computational and Graphical Statistics*, vol. 20, no. 2, pp. 316–334, Jan. 2011. DOI: 10.1198/jcgs.2011.09224. [Online]. Available: https://doi.org/10.1198/jcgs.2011.09224.

[108] M. Febrero-Bande and M. Oviedo de la Fuente, "Statistical computing in functional data analysis: The R package fda.usc," *Journal of Statistical Software*, vol. 51, no. 4, pp. 1–28, 2012. [Online]. Available: https://www.jstatsoft.org/v51/i04/.

[109] Q. F. Gronau, A. Sarafoglou, D. Matzke, *et al.*, "A tutorial on bridge sampling," *Journal of Mathematical Psychology*, vol. 81, pp. 80–97, Dec. 2017. DOI: 10.1016/j.jmp.2017.09.005. [Online]. Available: https://doi.org/10.1016/j.jmp.2017.09.005.

[110] N. Chénier, P. M. Magyar, L. Emmenegger, M. F. Lehmann, and J. Mohn, "Fractionation of nitrogen and oxygen isotopic composition in N2O produced by bacterial denitrification," May 2023. DOI: 10.5194/egusphere-egu23-5210. [Online]. Available: https://doi.org/10.5194/egusphere-egu23-5210.

[111] A. Vehtari, A. Gelman, and J. Gabry, "Practical bayesian model evaluation using leave-one-out cross-validation and waic," 2015. DOI: 10.48550/ARXIV.1507.04544. [Online]. Available: https://arxiv.org/abs/1507.04544.

[112] E. Stoll, E. Diaz-Pines, D. Reinthaler, *et al.*, "Understanding soil N2O emissions and production pathways in a changing climate by coupling automated chambers with isotope measurements&amp#160," Mar. 2022. DOI: 10.5194/egusphere-egu22-500. [Online]. Available: https://doi.org/10.5194/egusphere-egu22-500.